)LUME 108 PART B NUMBER 37 JANUARY 1961

The Proceedings

OF

THE INSTITUTION OF ELECTRICAL ENGINEERS

FOUNDED 1871: INCORPORATED BY ROYAL CHARTER 1921

PART B

ELECTRONIC AND COMMUNICATION ENGINEERING
(INCLUDING RADIO ENGINEERING)

Price Ten Shillings and Sixpence

THE INSTITUTION OF ELECTRICAL ENGINEERS

FOUNDED 1871 INCORPORATED BY ROYAL CHARTER 1921

PATRON: HER MAJESTY THE QUEEN

COUNCIL 1960–1961

President

SIR HAMISH D. MACLAREN, K.B.E., C.B., D.F.C.,* LL.D., B.Sc.

Past-Presidents

W. H. ECCLES, D.Sc., F.R.S.
THE RT. HON. THE EARL OF MOUNT EDGCUMBE, T.D.
J. M. DONALDSON, M.C.
PROF. E. W. MARCHANT, D.Sc.
H. T. YOUNG.
SIR GEORGE LEE, O.B.E., M.C.
J. R. BEARD, C.B.E., M.Sc.
SIR NOEL ASHBRIDGE, B.Sc.(Eng.).
SIR HARRY RAILING, D.Eng.
P. DUNSHEATH, C.B.E., M.A., D.Sc.(Eng.), LL.D.
SIR VINCENT Z. DE FERRANTI, M.C.
T. G. N. HALDANE, M.A.
PROF. E. B. MOULLIN, M.A., Sc.D., LL.D.
SIR ARCHIBALD J. GILL, B.Sc.(Eng.).
SIR JOHN HACKING.
COL. B. H. LEESON, C.B.E., T.D.
SIR HAROLD BISHOP, C.B.E., B.Sc.(Eng.), F.C.G.I.
SIR JOSIAH ECCLES, C.B.E., D.Sc.
THE RT. HON. THE LORD NELSON OF STAFFORD.
SIR W. GORDON RADLEY, K.C.B., C.B.E., Ph.D.(Eng.).
S. E. GOODALL, M.Sc.(Eng.), F.Q.M.C.
SIR WILLIS JACKSON, D.Sc., D.Eng., Le.D., F.R.S.

Vice-Presidents

B. DONKIN, B.A. O.W. HUMPHREYS, C.B.E., B.Sc. G. S. C. LUCAS, O.B.E., F.C.G.I. C. T. MELLING, C.B.E., M.Sc.Tech. A. H. MUMFORD, O.B.E., B.Sc.(Eng.).

Honorary Treasurer

C. E. STRONG, O.B.E., B.A., B.A.I.

Ordinary Members of Council

J. C. ARKLESS, B.Sc.
PROF. H. E. M. BARLOW, Ph.D., B.Sc.(Eng.).
D. A. BARRON, M.Sc.
C. O. BOYSE, B.Sc.(Eng.).
F. H. S. BROWN, C.B.E., B.Sc.
PROF. M. W. HUMPHREY DAVIES, M.Sc.
SIR JOHN DEAN, B.Sc.
L. DRUCQUER.
J. M. FERGUSON, B.Sc.(Eng.).
D. C. FLACK, B.Sc.(Eng.), Ph.D.
R. J. HALSEY, C.M.G., B.Sc.(Eng.) F.C.G.I.
R. A. HORE, M.A., B.Sc.
J. S. MCCULLOCH.
PROF. J. M. MEEK, D.Eng.
THE HON. H. G. NELSON, M.A.
H. V. PUGH.
J. R. RYLANDS, M.Sc., J.P.
R. L. SMITH-ROSE, C.B.E., D.Sc., Ph.D.
G. A. V. SOWTER, Ph.D., B.Sc.(Eng.).
H. G. TAYLOR, D.Sc.(Eng.).
D. H. TOMPSETT, B.Sc.(Eng.).

Chairmen and Past-Chairmen of Sections

Electronics and Communications:
T. B. D. TERRONI, B.Sc.
†M. J. L. PULLING, C.B.E., M.A.

Measurement and Control:
C. G. GARTON.
†PROF. A. TUSTIN, M.Sc.

Supply:
J. E. L. ROBINSON, M.Sc.
†J. R. MORTLOCK, Ph.D., B.Sc.(Eng.).

Utilization:
J. M. FERGUSON, B.Sc.(Eng.).
†T. E. HOUGHTON, M.Eng.

Chairmen and Past-Chairmen of Local Centres

East Midland Centre:
LT.-COL. W. E. GILL, T.D.
†D. H. PARRY, B.Sc.

Mersey and North Wales Centre:
D. A. PICKEN.
†T. A. P. COLLEDGE, B.Sc.(Eng.).

North-Eastern Centre:
D. H. THOMAS, M.Sc.Tech., B.Sc.(Eng.).
†H. WATSON-JONES, M.Eng.

North Midland Centre:
F. W. FLETCHER.
†PROF. G. W. CARTER, M.A.

North-Western Centre:
F. LINLEY.
†F. J. HUTCHINSON, M.Eng.

Northern Ireland Centre:
J. MCA. IRONS.
†T. S. WYLIE.

Scottish Centre:
R. B. ANDERSON.
†J. A. AKED, M.B.E.

South Midland Centre:
BRIGADIER F. JONES, C.B.E., M.Sc.
†G. F. PEIRSON.

Southern Centre:
R. GOFORD.
†W. D. MALLINSON, B.Sc.(Eng.).

Western Centre:
A. C. THIRTLE.
†H. JACKSON, B.Sc.(Eng.).

† Past Chairman.

ELECTRONICS AND COMMUNICATIONS SECTION COMMITTEE 1960–1961

Chairman

T. B. D. TERRONI, B.Sc.

Vice-Chairmen

R. J. HALSEY, C.M.G., B.Sc.(Eng.), F.C.G.I. J. A. RATCLIFFE, C.B.E., M.A., F.R.S.

Past-Chairmen

G. MILLINGTON, M.A., B.Sc. M. J. L. PULLING, C B.E., M.A.

Ordinary Members of Committee

W. H. ALDOUS, B.Sc., D.I.C.
D. A. BARRON, M.Sc.
P. A. T. BEVAN, C.B.E., B.Sc.
J. BROWN, M.A., Ph.D.
G. G. GOURIET.
J. MOIR.
L. J. I. NICKELS, B.Sc.(Eng.).
W. J. PERKINS.
N. C. ROLFE, B.Sc.(Eng.).
K. F. SANDER, M.A., Ph.D., B.Sc.
J. A. SAXTON, D.Sc., Ph.D.
T. R. SCOTT, D.F.C., B.Sc.
F. J. D. TAYLOR, O.B.E., B.Sc.(Eng.).
R. C. G. WILLIAMS, Ph.D., B.Sc.(Eng.).
R. C. WINTON, B.Sc.
A. J. YOUNG, B.Sc.(Eng.).

The President (*ex officio*).
The Chairman of the Papers Committee.
PROFESSOR H. E. M. BARLOW, Ph.D., B.Sc.(Eng.) (representing the Council).
E. H. COOKE-YARBOROUGH, M.A. (Co-opted Member).
E. D. TAYLOR, M.Sc. (representing the North-Eastern Measurement and Electronics Group).
H. V. BECK, B.Sc., M.A. (representing the Cambridge Electronics and Measurement Group).
S. D. MELLOR, B.Eng. (representing the North-Western Electronics and Communications Group).

And

J. R. POLLARD, M.A. (representing the East Midland Electronics and Control Group).
J. STEWART, M.A., B.Sc. (representing the Scottish Electronics and Measurement Group).
R. E. YOUNG, B.Sc.(Eng.) (representing the South Midland Electronics and Measurement Group).

The following nominees:
Royal Navy: CAPTAIN J. S. RAVEN, B.Sc., R.N.
Army: COL. R. G. MILLER, M.A.
Royal Air Force: GROUP CAPTAIN D. W. ROWSON, B.Sc.(Eng.), R.A.F.

MEASUREMENT AND CONTROL SECTION COMMITTEE 1960–1961

Chairman

C. G. GARTON.

Vice-Chairmen

W. S. ELLIOTT, M.A.; A. J. MADDOCK, D.Sc.

Past-Chairmen

PROFESSOR A. TUSTIN, M.Sc.; J. K. WEBB, M.Sc.(Eng.), B.Sc.Tech.

Ordinary Members of Committee

S. S. CARLISLE, M.Sc.
W. J. JEFFERSON.
C. A. LAWS.
A. C. LYNCH, M.A., B.Sc.
R. E. MARTIN.
A. NEMET, Dr.Sc.Techn.
S. N. POCOCK.
W. RENWICK, M.A., B.Sc.
G. F. TAGG, Ph.D., B.Sc.(Eng.).
R. D. TROTTER, B.Sc.(Eng.).
J. H. WESTCOTT, B.Sc.(Eng.), Ph.D.
F. C. WIDDIS, B.Sc.(Eng.), Ph.D.

The President (*ex officio*).
The Chairman of the Papers Committee.
G. A. V. SOWTER, Ph.D., B.Sc.(Eng.) (representing the Council).
C. C. BAXENDALE (representing the North-Eastern Measurement and Electronics Group).
A. CHORLTON, B.Sc.Tech. (representing the North-Western Measurement and Control Group).

And

H. M. GALE, B.Sc.(Eng.) (representing the South Midland Electronics and Measurement Group).
D. R. HARDY, Ph.D., M.Sc.(Eng.) (representing the East Midland Electronics and Control Group).
W. H. P. LESLIE, B.Sc. (representing the Scottish Electronics and Measurement Group).
D. L. A. BARBER, B.Sc.(Eng.) (nominated by the National Physical Laboratory).

Secretary

W. K. BRASHER, C.B.E., M.A., I.E.E.

Deputy Secretary

F. JERVIS SMITH, M.I.E.E.

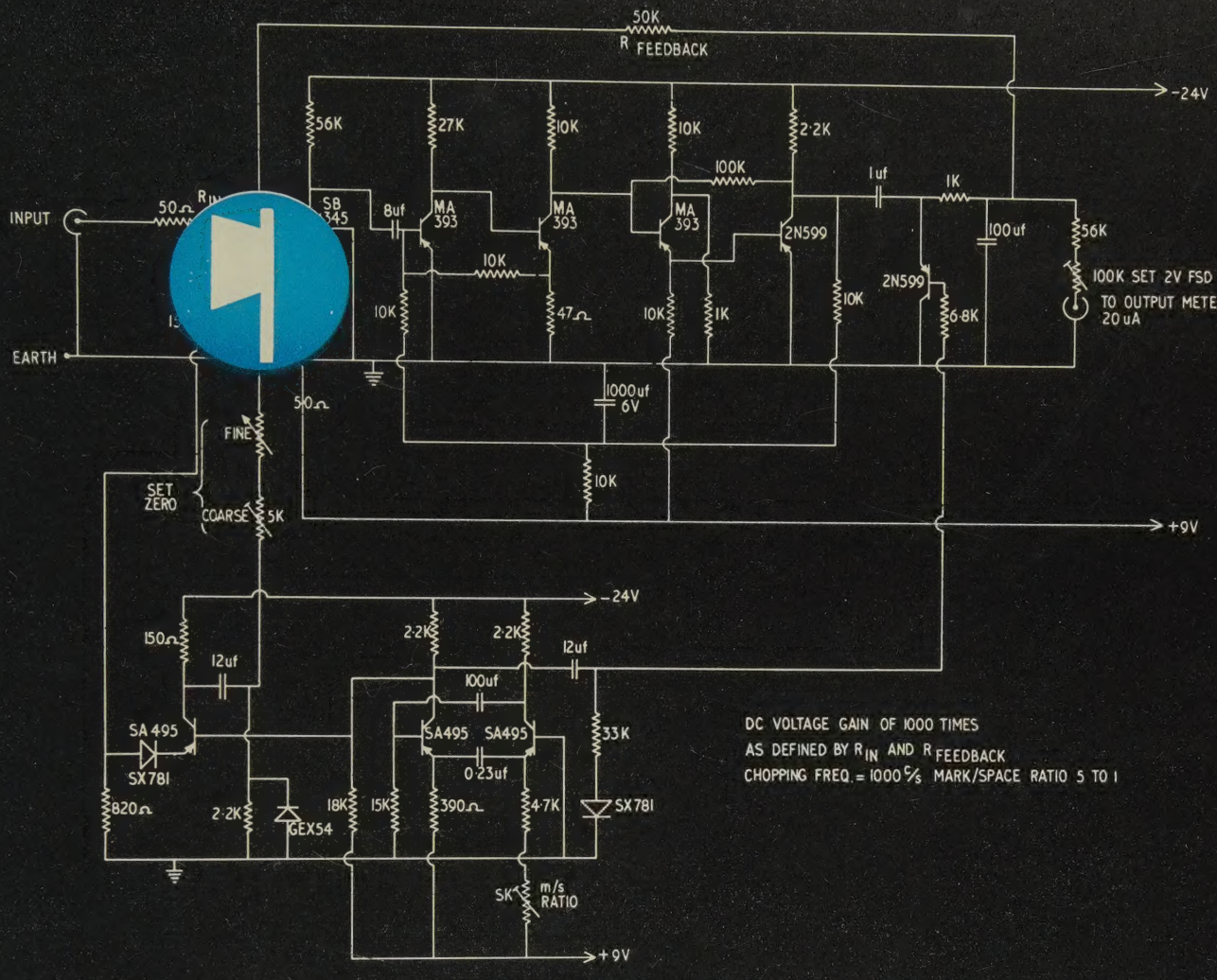
Principal Assistant Secretary

F. C. HARRIS.

Editor-in-Chief

G. E. WILLIAMS, B.Sc.(Eng.), M.I.E.E.

DC CHOPPER TYPE AMPLIFIER



SEMICONDUCTORS TRANSISTORS

FOR CHOPPER APPLICATIONS

Semiconductors' new Silicon Chopper Transistors are worth knowing about. The exclusive Semiconductors Limited Silicon Surface-Alloy process permits independent control of resistivity, base width and electrode diameter, allowing transistors to be made specifically for chopper applications.

For the first time in the United Kingdom transistors are in production with these key performance features :—

Offset voltage guaranteed to be below 2 millivolts or 5 millivolts according to type.

Very low leakage currents.

Low hole-storage factor.

Low collector capacitance.

Minimum f_T of 10 Mc/s — minimising transient problems and permitting wide band widths.

Write for Data Sheets to

Semiconductors Limited

Cheney Manor · Swindon · Wiltshire Tel: Swindon 6421

ONE OF THE **Plessey** GROUP OF COMPANIES

USE MULLARD TRIGGER TUBES

in control and switching circuits for . . .

Life and Reliability
Mullard cold cathode tubes offer a life expectancy and reliability comparable with the highest quality components used in electronic equipment.

High Sensitivity
Only a small energy pulse is required to "trigger" the main discharge. A gain of 100,000 may be achieved.

Instantaneous Operation
Cold cathode tubes are always ready for instantaneous operation. They are ideal for stand-by and intermittent duty applications.

The Mullard range at a glance

Z700U Low cost subminiature tube for use in computing/switching circuits.

Z700W Subminiature tube with two trigger electrodes for use in reversible ring counters.

Z701U Subminiature tube particularly suited for speech passing in electronic telephone exchanges.

Z803U Extremely stable, close tolerance miniature tube.

Z900T Miniature tube intended primarily for intermittent operation from d.c. or 115 volt a.c. supplies.

Full information on the range of Mullard trigger tubes is available —please write or telephone Mullard House

MULLARD LIMITED
Mullard House · Torrington Place
London W.C.1 · Tel: LANgham 6633

Mullard
GOVERNMENT AND INDUSTRIAL VALVE DIVISION

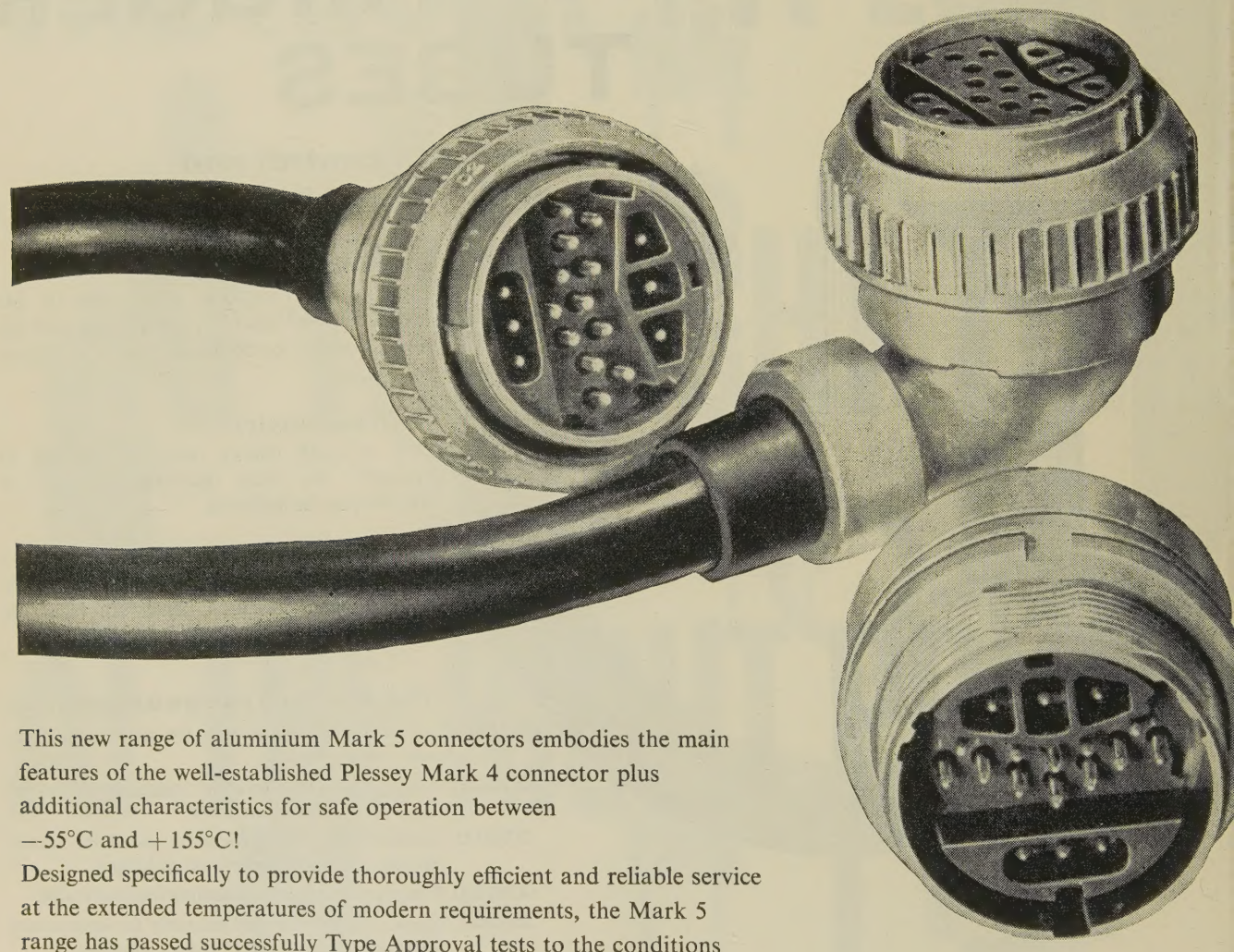
Mullard

MVT402a

from -55°C to +155°C

Plessey mark 5

connectors



This new range of aluminium Mark 5 connectors embodies the main features of the well-established Plessey Mark 4 connector plus additional characteristics for safe operation between —55°C and +155°C!

Designed specifically to provide thoroughly efficient and reliable service at the extended temperatures of modern requirements, the Mark 5 range has passed successfully Type Approval tests to the conditions specified in DEF 5321 (July 1958), maintaining a pressure sealing of 20 lb. p.s.i. between —40°C and +155°C. Where pressure sealing is not essential, efficient operation to —55°C is attained.

This standard of performance has resulted in the Mark 5 Connector being adopted by the Ministry of Aviation as the Pattern 104 connector.

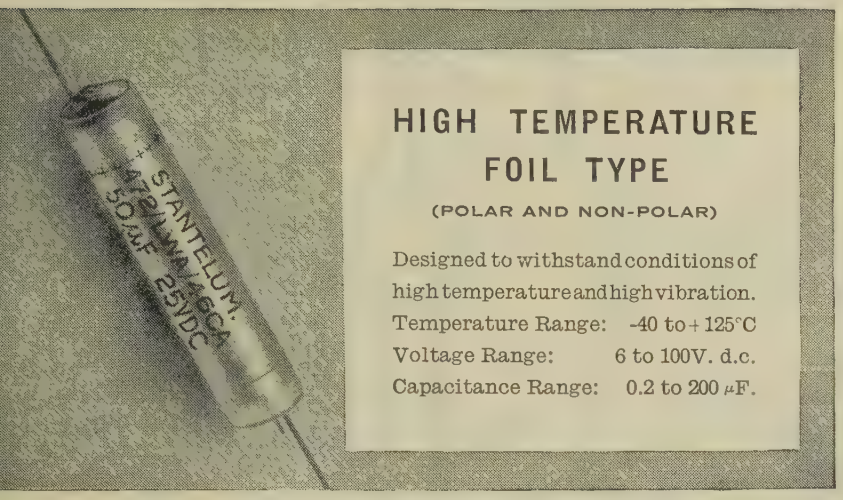
Wiring and Connectors Division

THE PLESSEY COMPANY LIMITED · Cheney Manor · Swindon · Wilts · Swindon 6251

Overseas Sales Organisation: PLESSEY INTERNATIONAL LIMITED · ILFORD · ESSEX · ILFORD 3040

CV

STANTELUM



HIGH TEMPERATURE FOIL TYPE

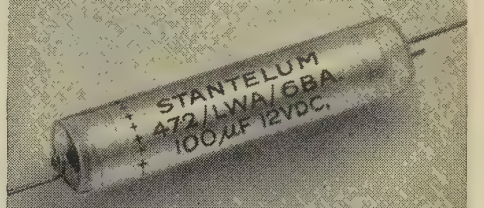
(POLAR AND NON-POLAR)

Designed to withstand conditions of high temperature and high vibration.
Temperature Range: -40 to +125°C
Voltage Range: 6 to 100V. d.c.
Capacitance Range: 0.2 to 200 μF.

STANDARD FOIL TYPE

(POLAR AND NON-POLAR)

Type approved to R.C.S. 134B
Temperature Range: -40 to + 85°C.
Voltage Range: 6 to 150V. d.c.
Capacitance Range: 0.15 to 200 μF.



TANTALUM electrolytic capacitors

STC manufacture a full range of tantalum electrolytic capacitors available in the following types:

MINIATURE FOIL TYPE

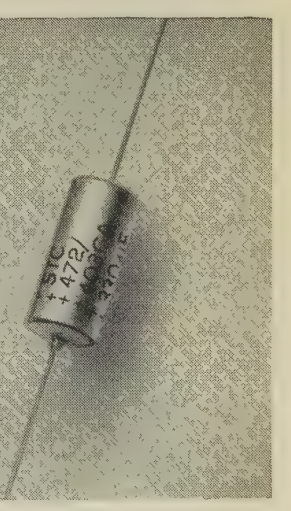
(POLAR)

A foil type tantalum capacitor in its most economical form.
Temperature Range: -25 to +70°C.
Voltage Range: 3 to 25V. d.c.
Capacitance Range: 1.5 to 16 μF.

SOLID TYPE

(POLAR)

Sintered Slug and solid electrolyte construction.
Temperature Range: -55 to +85°C.
Voltage Range: 6 to 35V. d.c.
Capacitance Range: 1 to 220 μF.



Write for details of the above Capacitors to:

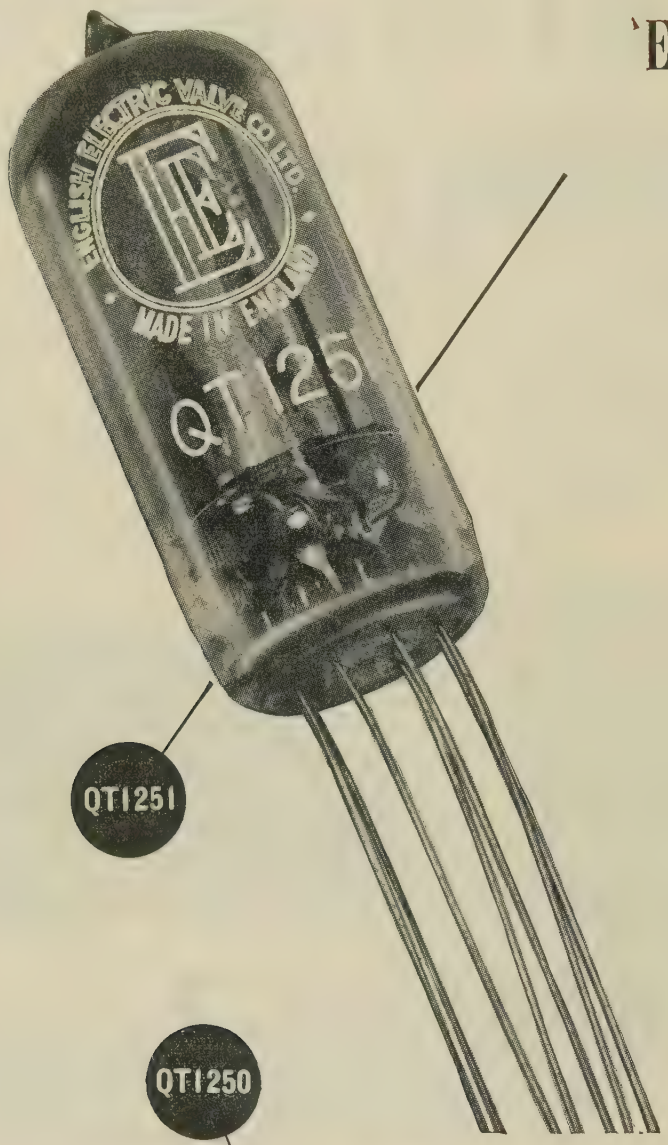
STC COMPONENTS GROUP

Standard Telephones and Cables Limited

Registered Office: Connaught House, Aldwych, London W.C.2.

CAPACITOR DIVISION: BRIXHAM ROAD · PAIGNTON · DEVON

60/7MC



'ENGLISH ELECTRIC'

TRIGGER TUBES

QT1251 was developed by the ENGLISH ELECTRIC VALVE COMPANY LTD from the well known 5823 with the object of reducing the tolerance on trigger breakdown voltage. Flying leads have been incorporated for direct wiring-in. The improved characteristics find particular application in electronic telephone exchanges.

A B7G version of the QT1251 is available with the type number QT1250.

Abridged data is given here, for full information on these tubes please write to the address below.

Characteristics	QT1250 & QT1251	5823
Peak Cathode Current max. (mA)	100	100
Mean Cathode Current max. (mA)	25	25
Operating Voltage max. (V)	68	85
Trigger Breakdown Voltage max.(V)	80	105
*Transfer Current nom. (μA)	140	135

**at Va=100V*

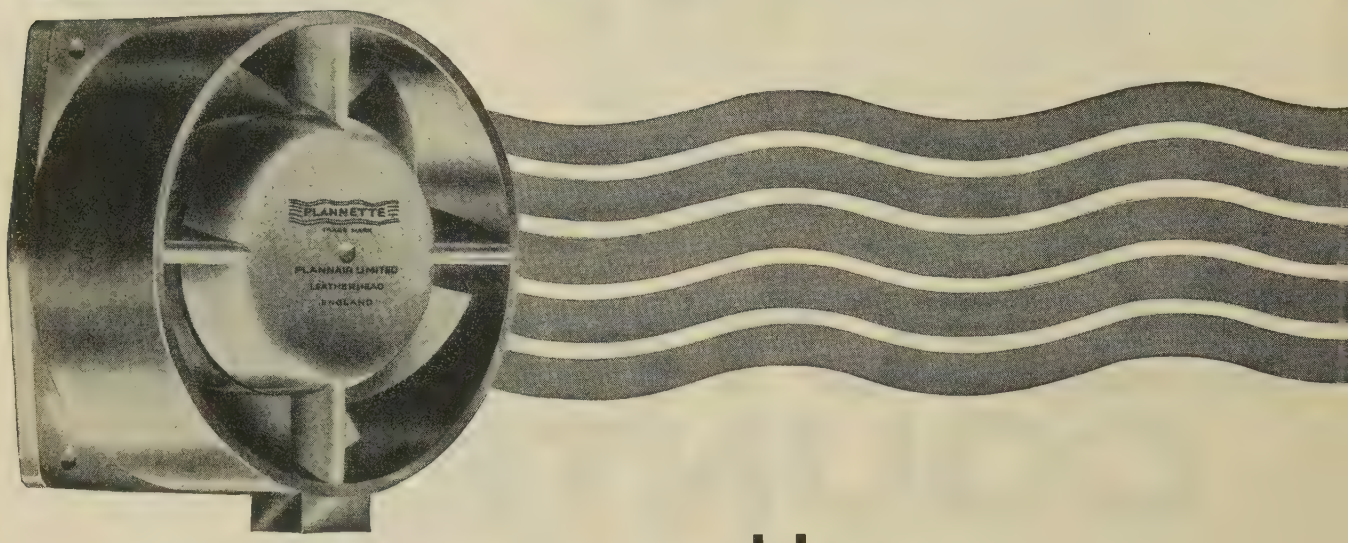
AGENTS THROUGHOUT THE WORLD

ENGLISH ELECTRIC VALVE COMPANY LTD

CHELMSFORD · ENGLAND Telephone: CHELMSFORD 3491

AP 163

meet the PLANNETTE



—it's only 2″ deep

...a new blower for forced air cooling in confined spaces

Here, for the first time, is a blower that makes efficient forced air cooling in confined spaces a practical proposition. In the new PLANNETTE the entire fan assembly is only 2″ deep and this makes it ideal for projects such as computers, electronic black boxes, mobile radios, or electronic instruments. The PLANNETTE can be mounted either inside or on top of the cabinet and thus it makes cooling possible where space is insufficient to fit a conventional blower. Its performance is quite exceptional. Write now for detailed leaflet.

DESCRIPTION

The new 'Plannette' fans are available in two sizes, 4½″ and 6″ diameter. The casing is light alloy, die cast, and the impeller blades, which are of a special aerofoil section, are moulded plastic.

The motors are a.c. and may be arranged either for 230V 1-ph. 50 cycles or 110V 1-ph. 50 cycles. The nominal operating speed is 2,700 r.p.m. Noise level is extremely low.

PLANNAIR

PLANNAIR LIMITED

Windfield House • Leatherhead • Surrey
Tel: Leatherhead 4091/3, 2231

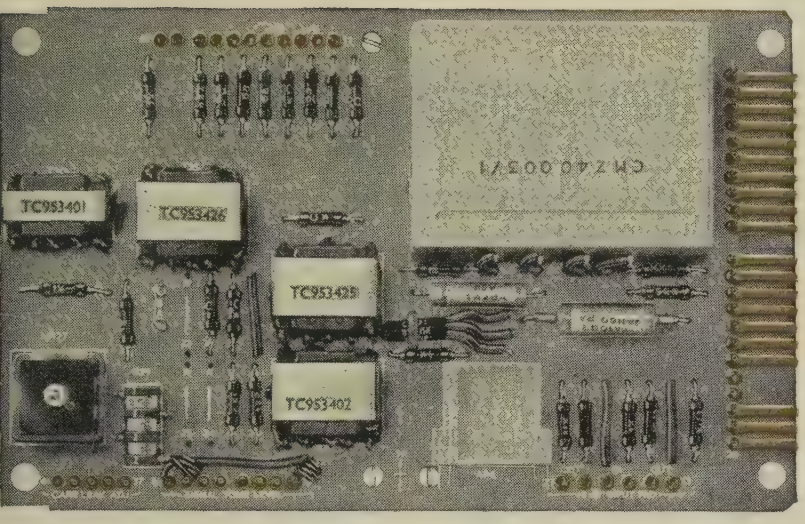
PLA

For Your Next Project

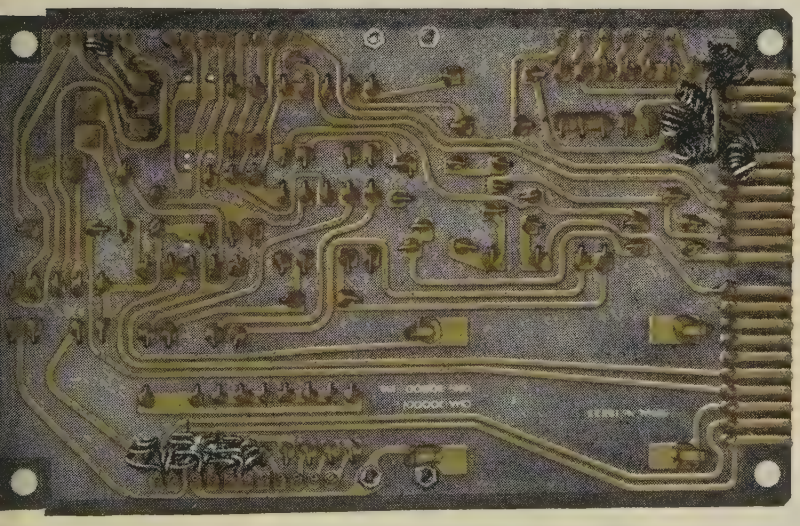
ATE

Transmission Equipment

TYPE CM



Front and back of channel equipment board, half actual size.



THESE PHOTOGRAPHS of a typical Type CM sub assembly show that printed wiring, transistors, ferrites, tantalum capacitors and gold-plated contacts are just some of the more obvious features of the compact equipment design.

To the user, the fact that every feature has been submitted to the most prolonged and painstaking proving before production is equally important and an assurance that maintenance and installation requirements have been minimised.

A.T.E. TRANSMISSION EQUIPMENT TYPE CM FOR LINE, CABLE AND RADIO SYSTEMS

For further information on Type CM Equipment write for Bulletin TEB 3001 to:—

AUTOMATIC TELEPHONE & ELECTRIC CO. LTD.

STROWGER HOUSE, ARUNDEL STREET, LONDON, W.C.2

TELEPHONE: TEMPLE BAR 9262

AT8951

These new STC complete packaged oscillator units represent a revolutionary step forward in crystal design. They are extremely small units for use as laboratory references or as the oscillator section of high performance equipment. The ruggedness of design together with the small size makes possible a degree of portability hitherto unknown in this type of unit. Standard frequency is 5 Mc/s.

Write for Technical Data Sheets.

1 in 10^9 per day frequency standard for laboratories at £195

One of a new, low priced range of frequency standards from STC.

- **Rugged**
- **Compact**
- **Light-weight**
- **Reliable**
- **Low Cost**

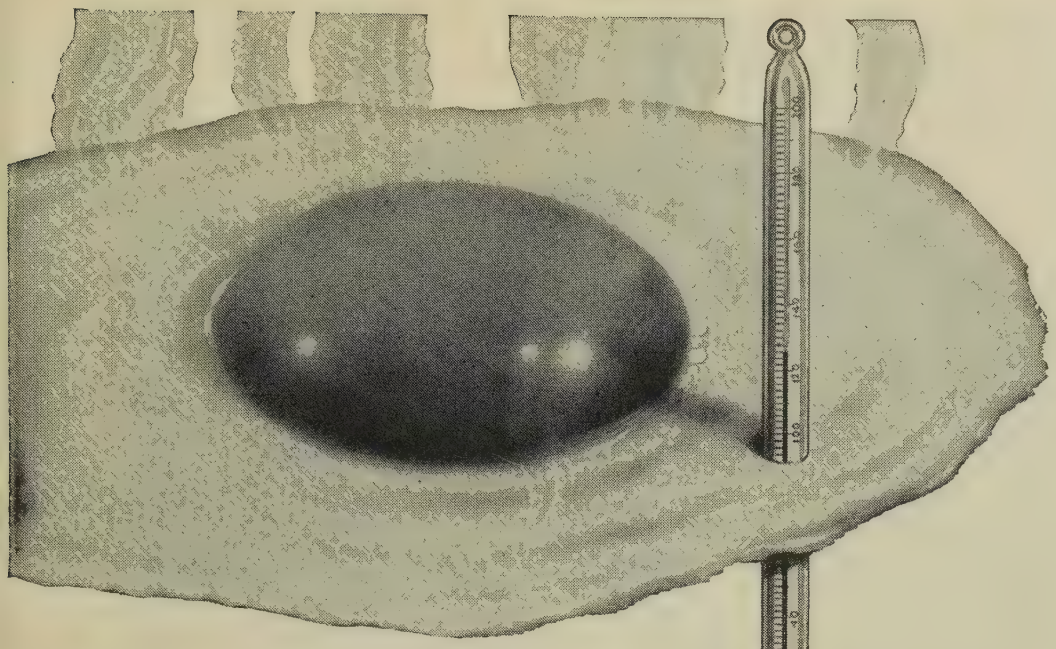
STC COMPONENTS GROUP

60/1 MQ

Standard Telephones and Cables Limited

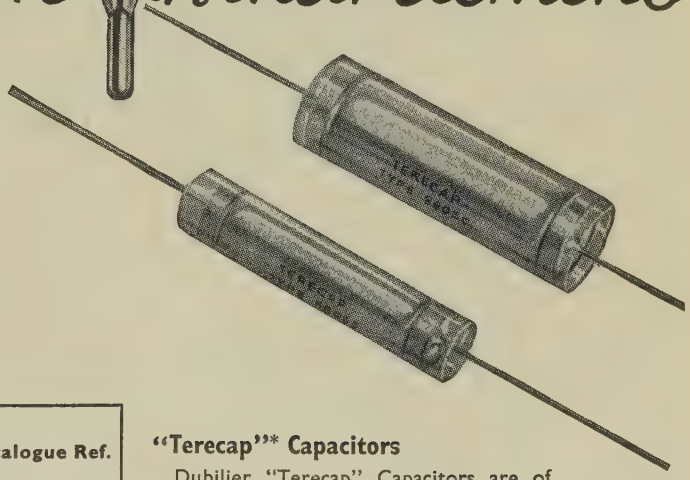
Registered Office: Connaught House, Aldwych, W.C.2.

QUARTZ CRYSTAL DIVISION: HARLOW · ESSEX



When it's hot enough to fry an egg

"Terecaps" are in their element



SIZES AND RATINGS

Capacitance	Working Voltage d.c. at 71°C	Working Voltage d.c. at 125°C	Test Volts d.c.	Dimensions ins. Length	Dimensions ins. Dia.	Catalogue Ref.
0.1	150	125	300	1¼	½	8801.C
0.25	150	125	300	1⅞	½	S-8803.C
0.5	150	125	300	1⅜	⅝	S-8800.C
1.0	150	125	300	1⅞	¾	S-8804.C
0.1	250	180	500	1¼	½	8801.C
0.25	250	180	500	2⅛	½	8803.C
1.0	250	180	500	2⅛	¾	8804.C
0.1	350	250	700	1⅝	½	8802.C
0.25	350	250	700	1⅞	¾	S-8804.C
1.0	350	250	700	2⅛	1	8806.C

Other capacitance values can be supplied to order. We invite your enquiries

"Terecap"* Capacitors

Dubilier "Terecap" Capacitors are of tubular form with extended foil metal electrodes fitted with wire tail terminations and incorporate a non-hygroscopic film dielectric. Being designed to meet abnormal atmospheric conditions such as obtain in tropical zones the capacitors are supplied hermetically sealed in metal containers with ceramic end-seals.

* *A Registered Dubilier Trade Mark.*

Dubilier "Terecap" Capacitors have these outstanding features:—

1. Can be used up to 125°C. with voltage de-rating above 70°C.
2. High insulation resistance, more than twenty times that of paper dielectric capacitors. (10,000ΩF at 20°C).
3. Compactness.
4. Excellent capacitance stability over a wide temperature range. (Normal capacitance tolerance ± 20%)
5. Power Factor 0.5% at 20°C. for 1 kc/s.

DUBILIER

DUBILIER CONDENSER CO. (1925) LTD., DUCON WORKS, NORTH ACTON, LONDON, W.3

As simple as

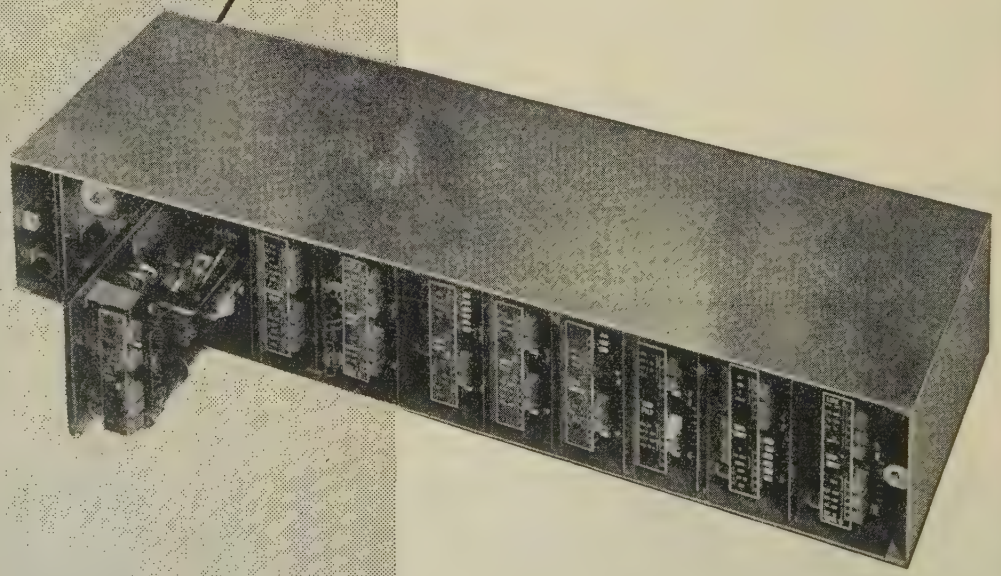
. . . plugging in the light!

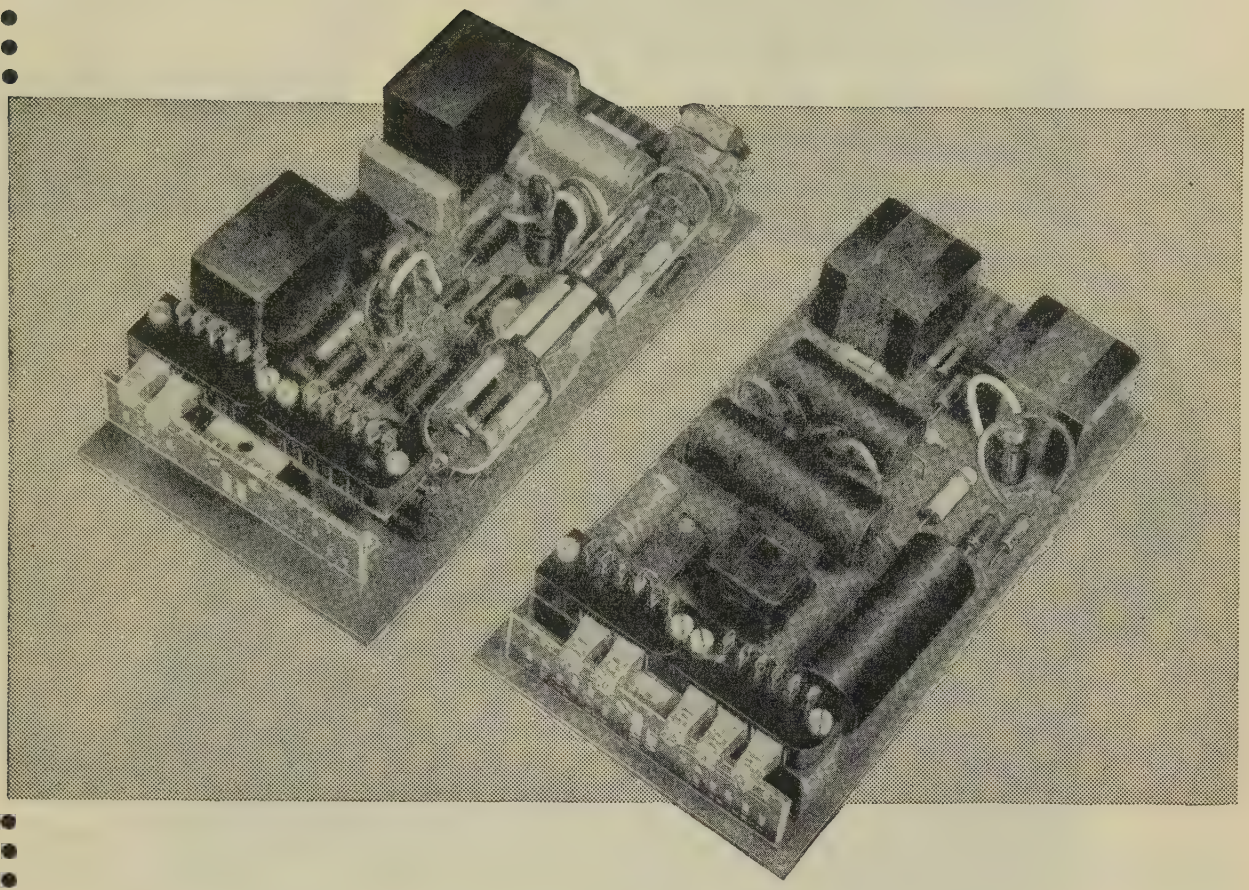
THE ERICSSON RURAL CARRIER TELEPHONE SYSTEM TYPE R.C.101

has been designed to combine ease of maintenance with small physical size and low power consumption. It is particularly suitable for use in remote areas where personnel trained to service carrier equipment may be limited. A fault may be cleared by substituting a spare for each plug-in unit in turn until the service is restored.

The system will provide up to ten additional speech circuits on an open wire line. The circuits are stackable and thus extra circuits may be added as demand increases. This feature together with a wide range of pole mounted 'drop-off' filter units offers a high degree of system flexibility. Channel re-allocation is easily achieved with this system of plug-in sub-units.

Illustrated on the opposite page are two typical plug-in units, a channel transmit unit and a hybrid unit with V.F. amplifier and limiter. Below is a single channel terminal with mains power unit and compandor units fitted.





brief specification

Audio bandwidth	300 c/s—2700 c/s
Overall frequency range (10 channels)	12 kc/s—172 kc/s
Carrier level to line (per channel)	+ 3 dbm
Minimum acceptable receive level from line	— 40 dbm
2 wire circuit equivalent	— 3 db
POWER CONSUMPTION:—	
Standby	10 m.a. at 12 v D.C.
During call	40 m.a. at 12 v D.C.

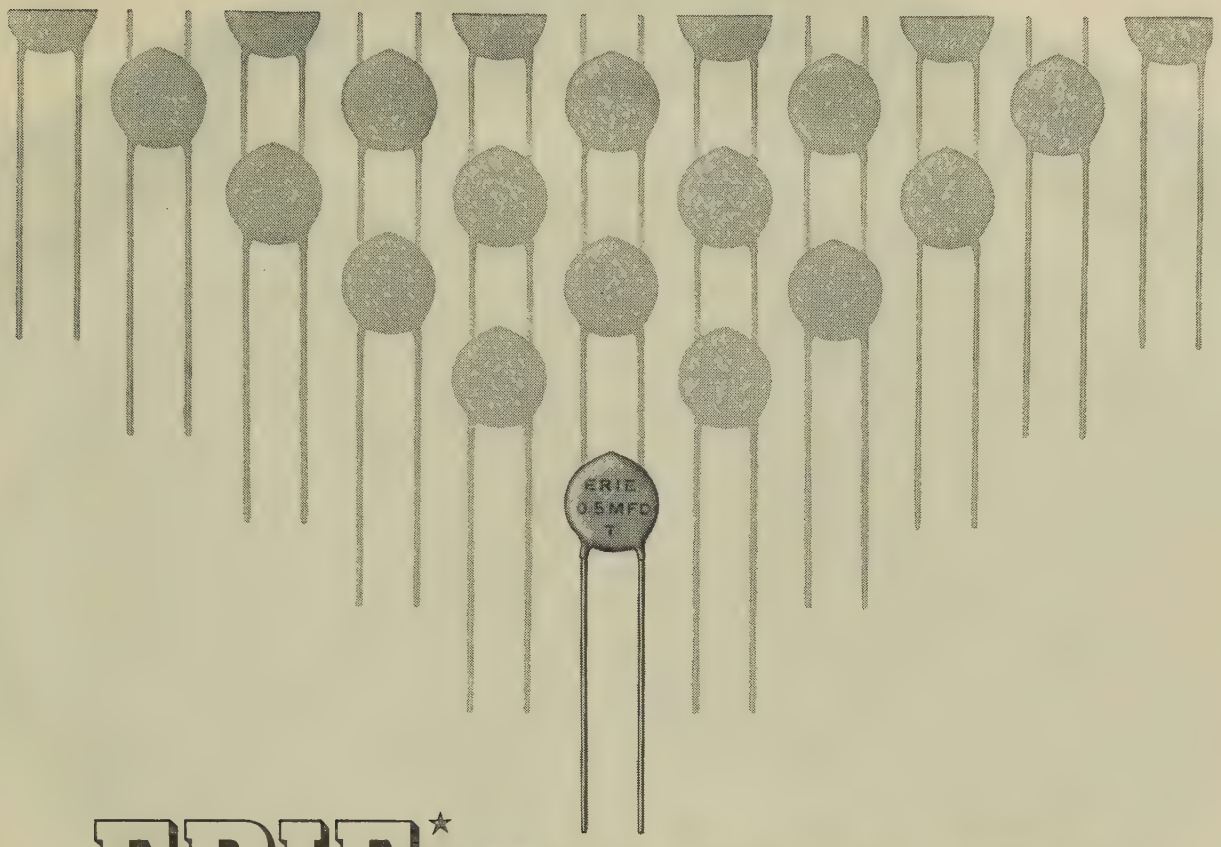
other features

- ★ **LOWER POWER CONSUMPTION.**
- ★ **BATTERY OR OPTIONAL MAINS OPERATION.**
- ★ **TRANSMITTED CARRIER FOR AUTOMATIC GAIN CONTROL AND SIGNALLING.**
- ★ **INTERCHANGEABLE SIGNALLING UNITS.**
- ★ **TRANSISTOR CIRCUITS.**
- ★ **COMPANDOR UNITS OPTIONAL.**
- ★ **AMPLE TEST POINTS.**
- ★ **ONE 10 CHANNEL TERMINAL ACCOMMODATED ON A WALL MOUNTING RACK 3′ 3″ HIGH.**
- ★ **SIMPLE INSTALLATION.**
- ★ **PORTABLE TEST SET AVAILABLE.**

For further information please write to:—

ERICSSON TELEPHONES LIMITED—ETELCO LIMITED

Head Office: 22 Lincoln's Inn Fields, London W.C.2. Tel: HOL. 6936 · Works: Beeston, Nottingham and Sunderland



ERIE★ Transcaps

Patent applied for

for all Transistor Circuits

SPECIFICATION

Diameter : 0.594 inches maximum

Thickness : 0.156 inches maximum

Capacitance : 0.5 mfd

Tolerance : −20% +50%

Working voltage : 3 volts d.c.

Power factor : Not greater than 5%, when measured at 1 kc/s, and less than 0.5 volts

Leakage resistance : Not less than 100,000 ohms, when measured at 3 volts

In line with the Erie policy of anticipating the component requirements of the future, the Erie Transcap capacitor is now added to our ever-increasing range of components for use with transistors.

Designed specifically as a small, reliable, high capacitance, low voltage, coupling, and by-pass capacitor, the Erie developed Transcap is manufactured entirely at our Great Yarmouth factory.

Style T, shown here in its actual physical size, is but a forerunner of the wide range, in differing values and voltages, which will ultimately emerge.

ERIE★ RESISTOR LIMITED

I, HEDDON STREET, LONDON, W.1

Telephone: REGent 6432

FACTORIES

Great Yarmouth and Tunbridge Wells, England: Trenton, Ont., Canada: Erie, Pa., Holly Springs, Miss., and Hawthorne, Cal., U.S.A.

★*Registered Trade Mark*

Connects Seattle to Ketchikan (Alaska).

Extending from San Francisco to Hawaii.

From Oban (Scotland) to Clarenville (Newfoundland) extending to Sydney Mines (Nova Scotia). This system is known as TAT.

From Oban to Cornerbrook (Newfoundland), known as CANTAT.

Begins at Gairloch on the west coast of Scotland and extends to Torshavn (Faroes). Thence to Reykjavik (Iceland), known as SCOTICE.

Linking Vestmannaeyjar in Iceland with Frederiksdal (Greenland) and Cornerbrook, Newfoundland — called ICECAN.

Starts at Manahawkin (U.S.A.) terminating at Bermuda.

SELLING AGENTS

Australia and New Zealand:
Telephone Manufacturing Company (A'sia) Proprietary Limited, Sydney, New South Wales.

Canada and U.S.A.:
Telephone Manufacturing Company Limited, Toronto, Ontario.

All other Countries (for transmission equipment only):
Automatic Telephone and Electric Company Limited, London.

now available to British Industry

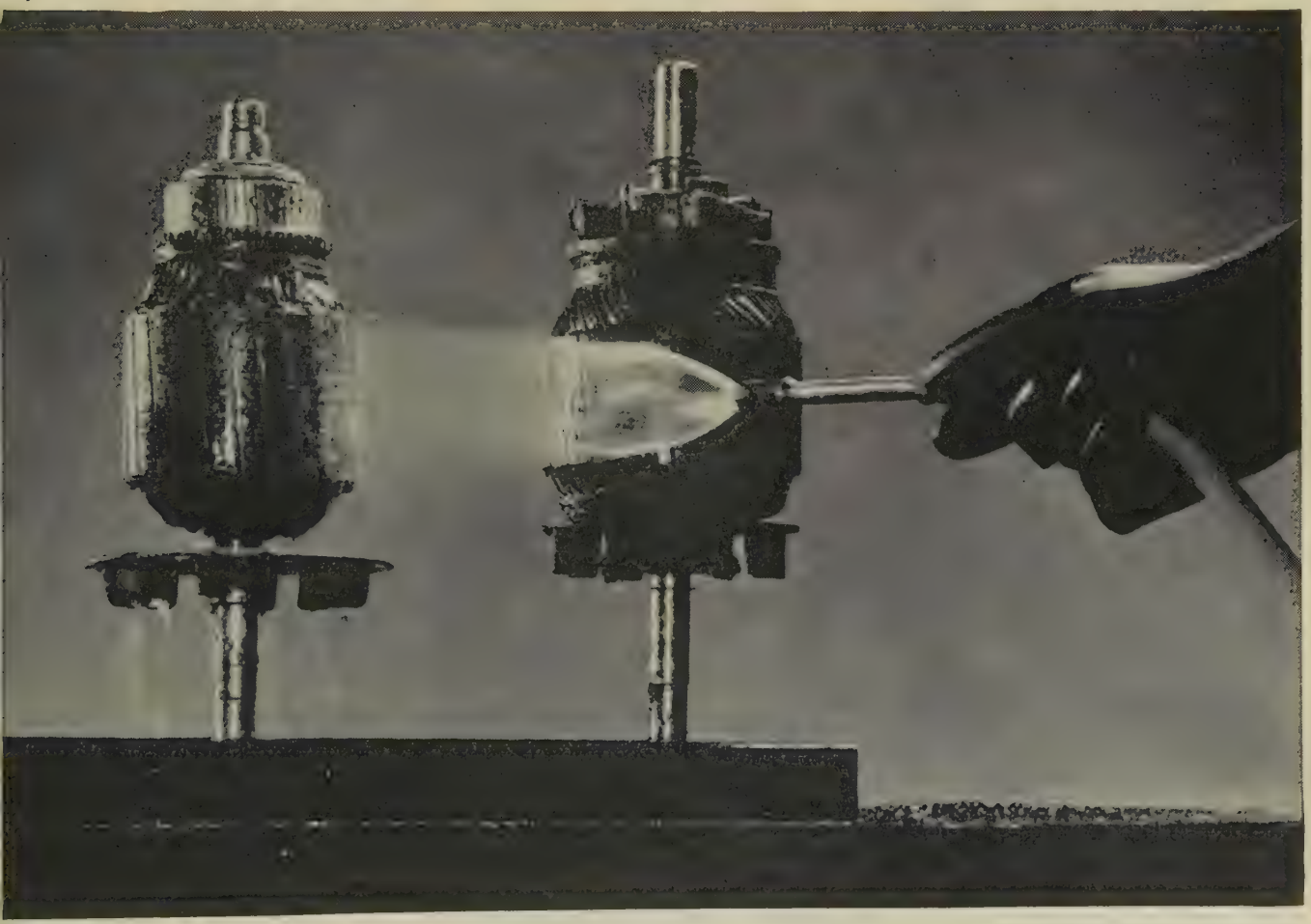
CHLOROTHENE NU

COLD CLEANS ELECTRICAL EQUIPMENT WITHOUT CORROSION

Safe, superior Chlorothene* NU, now revolutionizing cold cleaning operations in the U.S. electrical industry, eliminates corrosion of sensitive metals *and* insulation materials. Its exceptional stability (specially inhibited 1,1,1-trichloroethane) makes it ideal for cleaning even aluminium, zinc and white metal alloys, as well as plastics, rubbers and other materials. In fact, Chlorothene NU cleans everything from electric circuits to complex electronic brains. It has successfully cleaned finishes to which ultrasonics have proved injurious. **Safe in two ways!** Chlorothene NU greatly reduces the twin hazards of toxicity and fire. Its maximum allowable vapour concentration is 500 parts per million—comparable with many flammable solvents and substantially higher than most chlorinated solvents. Chlorothene NU practically eliminates the fire risks of flammable solvents because it has *no fire or flash point!*

Chlorothene NU cleans by dip, spray, slush or wipe methods —on the production line or for maintenance. It quickly removes greases, oils, waxes, tars and other soils, and dries fast without leaving a residue. For complete information about its specific advantages in the cleaning of electrical equipment, contact the distributor *NOW*.

**Trademark of The Dow Chemical Company, U.S.A.*



Distributor:

PENETONE-PARIPAN LIMITED

P.O. Box No. 10, Egham, Surrey Tel: Egham 3811 (10 lines)

DOW CHEMICAL COMPANY (U.K.) LIMITED

48 Charles Street, London, W.1. Tel: GROsvenor 5406

CHEMICALS
PLASTICS

DOW

TEXTILE FIBRES
PACKAGING FILMS

G.E.C. Variable Capacitance Diodes now available

The SVC11-17 series of silicon variable capacitance diodes is designed for use in parametric amplifiers at frequencies up to S-band. These devices are the first of their kind to be produced in this country. The diodes are suitable for use in radar and other communications systems and are designed for high power dissipation so that no elaborate precautions against surges are necessary.

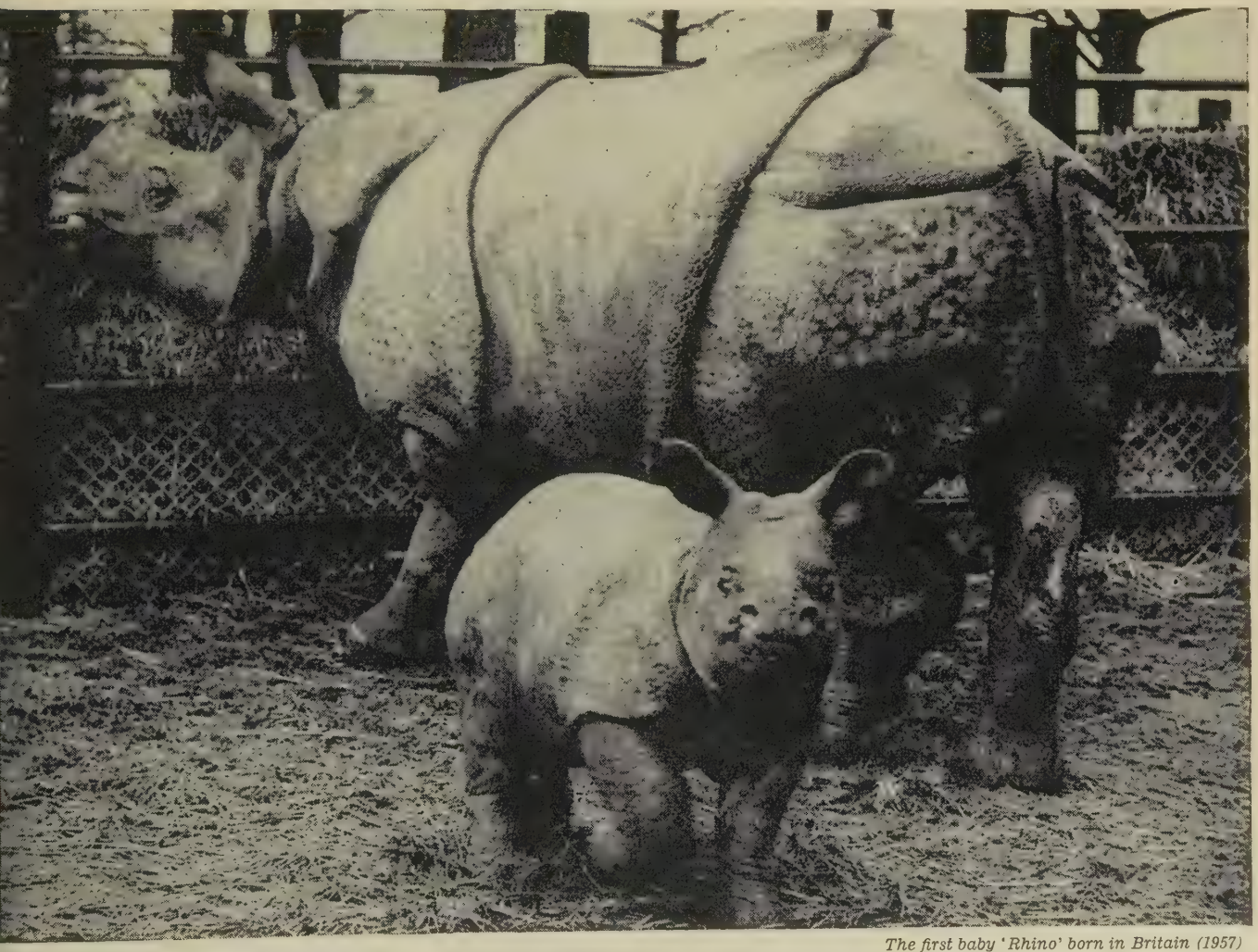
The device can be used as a frequency multiplier—an important step towards the realisation of a microwave receiver based entirely on solid state devices—and because of its very low forward impedance and very high reverse impedance it can also be used as a microwave switch.

The diode is mounted in a coaxial structure for direct insertion into coaxial and waveguide circuits and it has the low series inductance of 0.5 mμH. The range consists of diodes of minimum cut-off frequencies from 25 to 85 kMc/s in 10 kMc/s steps. Reverse polarity versions of SVC11 and 12 are SVC21 and 22 for use in balanced circuits.

For full information on this series please write to the address below

ISIT US AT THE PHYSICAL SOCIETY
XHIBITION STAND 130,
NEW HORTICULTURAL HALL, JANUARY 16 to 20.

Produced in Britain for the first time

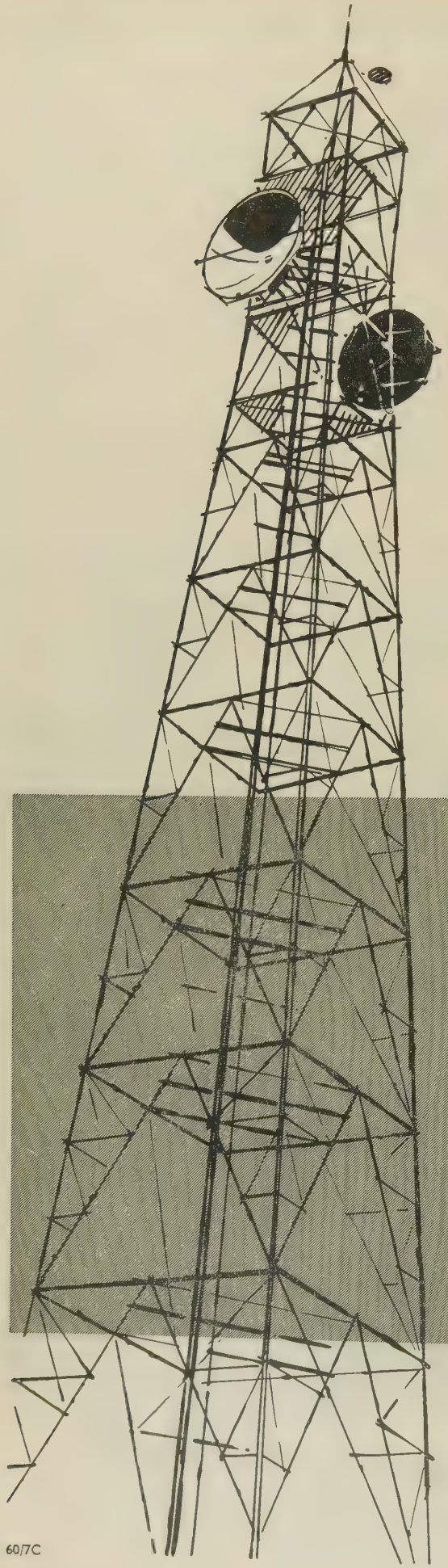


The first baby 'Rhino' born in Britain (1957)

G.E.C.

SEMICONDUCTORS

The General Electric Co. Ltd., Semiconductor Division,
School Street, Hazel Grove, Stockport, Cheshire.
(London Sales Office: TEMple Bar 8000, Ext. 10)



STC

PLAN, DESIGN, MANUFACTURE AND INSTALL COMPLETE S.H.F. SYSTEMS FOR TELEPHONE AND TELEVISION TRANSMISSION

STC can supply microwave radio systems complete with the necessary cable extensions and frequency division multiplex equipment.

STC EXPERIENCE

STC have supplied 4000 Mc/s systems having a capacity of over 4½ MILLION Telephone circuit miles and over 5000 Television channel miles.

STC are supplying systems to 18 countries

4000 Mc/s Terminal equipment cubicles.

COMMUNICATIONS GROUP

Experienced STC engineers are ready to help solve your communications problems.

Standard Telephones and Cables Limited

Registered Office: Connaught House, Aldwych, London, W.C.2

TRANSMISSION DIVISION: NORTH WOOLWICH · LONDON · E.16

60/7C

A new peak of performance

The new E.M.I. OSCILLOSCOPE TYPE WM16 is a wide-band, highly sensitive instrument of great flexibility and range. It has outstandingly high performance, yet is considerably cheaper than others in the same class. Compact and convenient to operate, the WM16 makes use of plug-in units to increase its flexibility and ease of maintenance. This is an ideal instrument for government establishments, universities and industrial concerns. In addition to meeting the needs of general laboratory electronic work, it can be used for radar, television, computers and millimicrosecond oscillography.

MAIN CHARACTERISTICS of the E.M.I. Oscilloscope type WM16

BAND WIDTH	DC – 40 Mc/s
MAXIMUM SENSITIVITY	50 mV/cm at full bandwidth
TIME BASE SWEEP SPEED	20 mμ secs/cm – 0.5 secs/cm
TIME BASE DELAY	1 μs/ – 0.15 secs
MEASURING ACCURACY	± 3%
SIGNAL DELAY	0.2 μ secs
PLUG-IN UNITS	These include: Wide band unit, Dual trace unit.
ANCILLARIES	Attenuator Probes, Viewing Mask, Termination Pad, Trolley, Cameras, Binding Post Adaptor

EMI

Please write for full details to :

E.M.I. ELECTRONICS LTD

INSTRUMENT DIVISION · HAYES · MIDDLESEX · SOUTHALL 2

INDEX OF ADVERTISERS

THE PROCEEDINGS OF
HE INSTITUTION OF ELECTRICAL ENGINEERS

EDITED UNDER THE SUPERINTENDENCE OF W. K. BRASHER, C.B.E., M.A., M.I.E.E., SECRETARY

DL. 108. PART B. NO. 37. JANUARY 1961

The Institution of Electrical Engineers
Paper No. 3355
Oct., 1960

INAUGURAL ADDRESS

By Sir HAMISH D. MACLAREN, K.B.E., C.B., D.F.C.,* LL.D., B.Sc., President.

'A REVIEW OF ELECTRICAL ENGINEERING IN THE ROYAL NAVY'

(*Address delivered before* THE INSTITUTION *6th October*, 1960.)

t is traditional and appropriate that your incoming President uld begin his Inaugural Address with an acknowledgment of high honour which the members of The Institution have con-red on him by electing him to that most eminent and onerous ll offices which an electrical engineer can fill—President of The titution. I am indeed deeply conscious of that honour, and ually conscious that I am not going to find it easy to maintain high standard which my predecessors in office have set.

am, however, tremendously encouraged, in contemplating coming year of office, by the insight gained during my five rs as a Vice-President of the service rendered to our Institution, l of the support given to the President, by all active members Headquarters and in the Provinces, and by our Secretary and staff. Also, as from the end of this month, I am fortunate eing spared the prospect of attempting to combine my duties President with those of the Director of Electrical Engineering the Admiralty. For this I am most grateful to the Board of miralty for having agreed to such a timely date for my retire-nt. I am also fortunate to have the prospect of the coming ilarating Presidential year to soften the blow of saying good-e to the Admiralty service in which for 34 years I have enjoyed rofessional life of absorbing interest.

There must be few facets of electrical engineering which are reflected in modern warship design, ranging from the genera-n and distribution of power by steam, Diesel, gas turbine and mic fission to the very complex installations required for nputation and communication for modern control systems, ed on thermionic, magnetic and semiconductor techniques. e fact that the cost of the electrical installation in a present-day rship approaches 30% of the cost of the ship as a whole will haps help to convey to you the importance of this aspect of ign and construction. Fortunately, the D.E.E. is not directly onsible for some of the more specialized applications: but is responsible for the installation of all electrical equipment l for its compatibility in operation in the ship as a whole.

n selecting a subject for this Address from such a wide field, ound myself suffering from an embarrassment of riches; but urity' came to my aid in eliminating many items, and in any case it seemed more appropriate to select a subject which would be of interest to the largest number of those good enough to honour me by their presence this evening.

I shall therefore, rather ambitiously, attempt to review the application of electrical engineering in the Royal Navy from the earliest days down to the present time.

1852 to 1904

Electricity seems to have been first applied to the service of the Royal Navy in 1852, when an electric telegraph was installed between Whitehall and the Commander-in-Chief, Plymouth. Up to then the 220 miles were spanned by 22 semaphore stations, and it was claimed that Whitehall could send a message and get an answer back in 15 minutes. This makes one long for the good old days! The first application of electricity afloat was in the 1870's, when electric firing of guns was introduced. Pile-type batteries were used consisting of 160 sheets of alternate copper and zinc, separated by 'fearnought'—a kind of flannel used for stokers' trousers—dipped in a mixture of vinegar, salt and water. These were later replaced by Leclanché cells, with an earth return, and this system has been retained to the present day in firing circuits. In 1875 the first dynamo was installed in a warship. This was to provide power for a searchlight for use against torpedo craft. It is probably this early association with torpedoes which led to the torpedo officer becoming responsible for all electrical apparatus afloat until 1946, when the Electrical Branch was founded. In 1876 searchlights were installed in H.M.S. *Minotaur*, using a Wilde generator, belt driven from a pumping engine. H.M.S. *Inflexible* was fitted with an electric lighting and searchlight installation in 1881.

In this ship Swan incandescent lamps were connected in series groups to an 800-volt d.c. generator, which is shown in Fig. 1. Needless to say, this was not very satisfactory, and within a year the first fatal accident is recorded, not altogether surprisingly; 80 volts was then adopted as standard, and by 1885 some 100 vessels had searchlights and 10 had electric lighting. The lamps, illustrated in Fig. 2, had carbon filaments which terminated in platinum wire loops fused into the glass bulb.

The cables were rubber insulated, with a serving of cotton tape

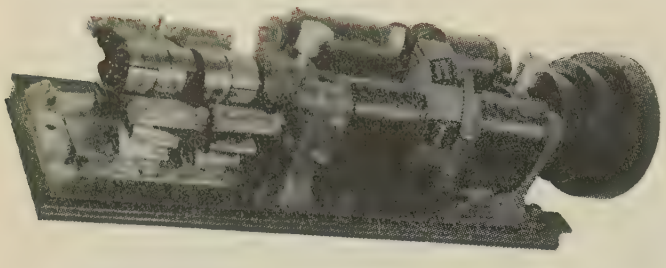


Fig. 1.—800-volt generator from H.M.S. *Inflexible*.
Overall length, 4 ft.

Fig. 2.—Early carbon-filament lamp.

coated with preservative. They were run in teak casings and embedded in putty. This was later replaced by a lead sheathing with layers of jute yarn as insulation instead of rubber. These all gave trouble, but a change to rubber insulation with a lead sheath in 1888 was much more successful, and this basic type of cable, with of course many improvements, continued to be used up to the end of the Second World War.

A first-class ship of 1886 was fitted with three dynamos each giving 200 amp at 80 volts. By 1900 the demand for electric power had grown to such an extent that dynamos giving 600 amp were being fitted. In 1899 the Admiralty Electrician, who had just been appointed to the Constructor Manager's Department of Portsmouth Dockyard, recommended that the voltage should be increased to 100, the voltage drop on an 80-volt system having become excessive as the electrical loading increased.

Although lighting and searchlights constituted the main loads in these early warships, the electric motor had been introduced and had proved its usefulness. By the end of the century, electric motors were used for ventilating fans, various auxiliary drives, capstans and even for training and elevating guns.

Electricity had now made a place for itself in warships, and in 1898 the Board of Admiralty decided that the torpedo officer was to be responsible at sea for electrical communication and electric motors. Soon after the Electrician had been appointed to Portsmouth Dockyard, the Board decided to have this new thing, 'electricity', thoroughly looked into, and they appointed a committee to do so. This resulted in the appointment in 1903 of Charles Wordingham—President I.E.E. in 1917 and 1918—as the first electrical engineer at Admiralty Headquarters. There was now recognition of electrical matters both at Headquarters and at sea.

In referring to this period, mention must be made of a very remarkable torpedo officer, Captain H. B. Jackson, who experimented with methods of detecting the newly discovered Hertzian waves. While serving in H.M.S. *Defiance* he developed a coherer and in 1896 he was successful in establishing W/T communication between H.M. Ships *Defiance* and *Scourge* over a range of 5800 yd. During that year he met Marconi and the two pursued research in the friendliest manner. In 1899 Jackson's persistence resulted in the Admiralty agreeing to Marconi installing his equipment in four ships with a communication range of up to 70 miles.

1905 to 1939

By 1905 it was being proposed to raise the voltage in warships to 220. Tests at the time demonstrated to the satisfaction of the Navy that '220-volt electric shocks were unpleasant but not dangerous'. They must have been tough in the Navy in those days! This voltage then became the standard for large ships. H.M.S. *Defence*, a cruiser completed in 1908, being the first ship to be so fitted. Hitherto there had been a single central switchboard such as that shown in Fig. 3. Each generator had

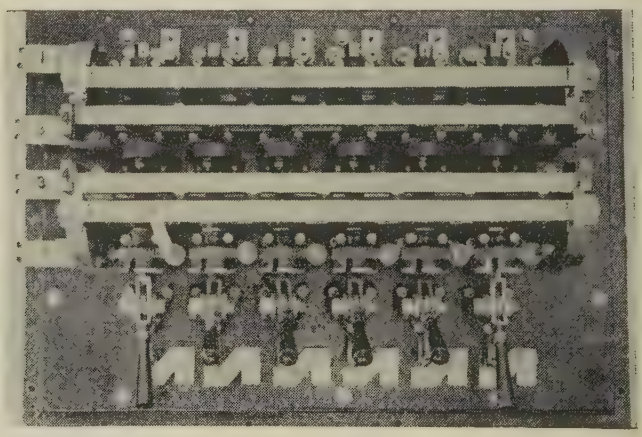


Fig. 3.—Early type of switchboard.

had its own set of busbars, and service feeders had been connected to alternative busbars by various types of change-over devices. The *Defence* was also the first ship to be fitted with a ring-main system of distribution, a system which, with the many improvements born of subsequent experience in peace and war, is still at sea in many of our larger ships. The main objective of the ring-main system as stated in the 1906 specification was 'to reduce the number of holes in vertical bulkheads and simplify the wiring for electric power and lighting generally, also to effect a saving in the amount of cable used'. It certainly went a long way towards achieving these objectives.

The ring-main system in the *Defence* was supplied by four generators, three of 105 kW and one of 210 kW capacity. Circuit

breakers were provided to protect each generator and at each service tapping point on the ring main. In the *Defence* the generators were grouped together, but in an improved design of ring-main system in H.M.S. *Orion*, in 1910, the four 200kW generators were installed in separate compartments.

In the early ring-main system, electrical equipment connected directly to it had to be in watertight enclosures capable of withstanding immersion in sea water to a specified depth. This was to ensure that flooding in any part of the ship involving the ring-main equipment would not lead to short-circuits and loss of all power. Hand-operated switches were provided to isolate damaged or flooded sections of the ring main, but the generators themselves, and the distribution equipment fed from the water-tight switchgear, were non-watertight.

Experience during action in the 1914–18 War revealed that the so-called watertight equipment rarely remained so in service, and that it was not easy to isolate damaged or flooded sections under the stress of action. During that war it was necessary to increase the number of isolating switches and to arrange for their operation from a more accessible deck and one which was not so susceptible to flooding. A system was also devised for automatically isolating non-watertight equipment from the ring main by means of a device called a 'main guard', shown in Fig. 4.

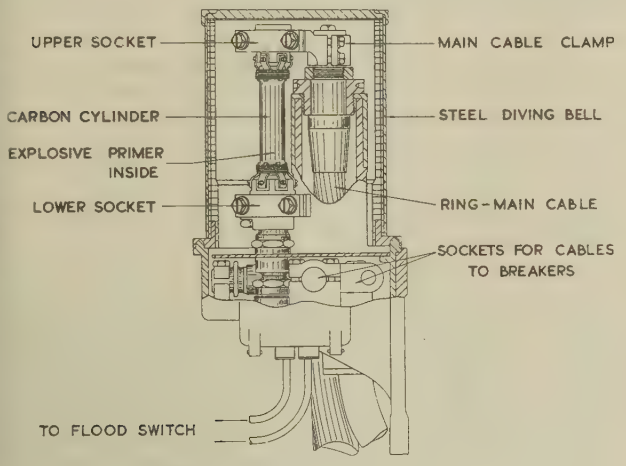


Fig. 4.—1 200 amp main guard.

This was an unusual piece of apparatus consisting of a carbon cylinder called an 'interrupter' enclosed within a miniature diving bell. The interrupter carried the full load of the generator or the group of branch breakers being supplied from the ring main. Explosive primers of the type used to initiate the firing of cordite in the ship's guns were mounted within the carbon interrupter and connected to electrodes in a flood switch, so that, when water in a compartment rose to the level of this switch, the resulting current fired the primers, shattering the carbon cylinder and thus disconnecting the item of non-watertight equipment from the ring-main before a short-circuit could occur.

The seriousness of flooding as a general problem is evidenced by the special measures adopted in H.M.S. *Hood*, which was designed in 1916 and, incidentally, continued in service till sunk by the *Bismark* in 1941. The *Hood* was fitted with eight 200kW d.c. generators, each having slip rings in addition to its commutator, so that an alternative a.c. supply of 200kVA at 135 volts 25c/s was obtainable when required from each generator to supply an a.c. ring main installed solely to provide power for submersible pumps to deal with flooded compartments.

After the 1914–18 War the Admiralty reviewed the experience to date of the generation and distribution of electrical power in warships, and in 1919 suggestions were called for from those concerned with the problem. The review included the possible introduction of alternating current.

The report issued in 1921 was strongly in favour of retaining the ring-main system in all except the smallest ships, but with certain modifications born of experience. It was also envisaged that the largest generator required in the future would be 250kW. The review justified the continuing use of direct current as having a balance of advantage compared with alternating current, having regard to the relatively restricted applications of electricity in warships at that time. Considerable importance, perhaps more than was justified, was attached to the ease with which speed control was readily available with direct current. It is interesting to note that one argument quoted against alternating current was that 'a Russian battleship, the *Volga*, had been fitted with an a.c. system which must have been a mistake because it was not repeated in later ships'.

Following this policy decision to retain direct current, effort was continued on developing the ring-main system, one notable improvement being the provision of automatic circuit-breakers, known as 'ring-main breakers', to enable sections of the ring main to be isolated. These operated on over-current and could be closed or opened from the main controlling switchboard. Another improvement was the development of the fuse-release switch to take the place of the main guard. This is toggle operated and held in the closed position against a spring by a stainless-steel fuse wire, which in turn is connected to the flood switch. The device, illustrated in Fig. 5, still survives in d.c. ships.

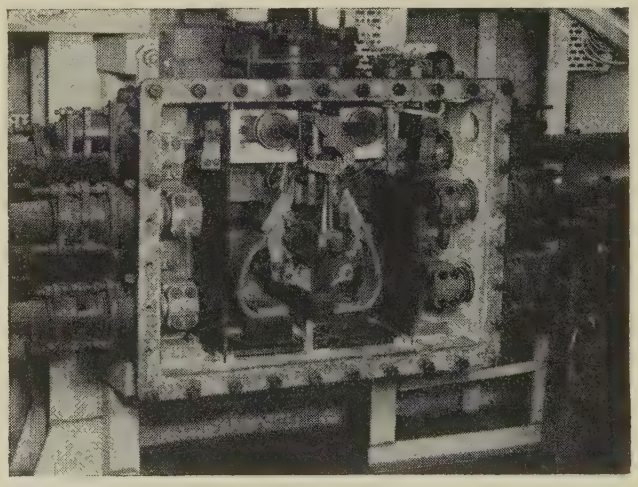


Fig. 5.—Fuse-release switch.

The possibilities of a change from direct current to alternating current were again seriously considered in 1932. It was assumed for the purpose of the investigation that a destroyer would have two 40kW turbo-generators with one standby set of 20kW, and a cruiser, four 225kW turbo-generators. These installations were still too small to justify a change to alternating current; but, had the very rapid rate of increase in demand for applications of electricity in the subsequent decade been foreseen, the decision might well have been different.

In the middle 1930's there was a considerable expansion in warship construction. New designs of cruisers and battleships were put in hand, and a new type of ship for the Royal Navy, the aircraft carrier, came into being. The first aircraft carriers were conversions from existing hulls, such as the *Furious* and the *Glorious*, but new ships designed from the outset as aircraft carriers were being constructed before 1939. These larger and

newer ships demanded more generating capacity, and the 1936 battleships, the *King George V* class, were designed with eight 300 kW generators, and the first aircraft carrier, the *Ark Royal*, had six 400 kW generators.

At this stage it was found necessary to take stock of the electrical supply position and deal with two important problems. First, there was growing doubt whether the fault protection system employed in the ring main was correct and reliable, particularly as regards its discriminating properties. It is essential that branch breakers should clear faults on feeder cables without the ring-main breakers opening, and similarly ring-main breakers should clear faults on the ring main itself without the generator supply breakers opening. Second, there was a lack of knowledge of the short-circuit current characteristics of the larger systems and of the performance of the circuit-breakers on low-impedance faults.

H.M.S. *London* was made available in late 1938 to enable an assessment of the problems to be made. The trials consisted of imposing a comprehensive series of low-impedance faults on the system with various combinations of generators running in parallel in order to determine the maximum possible short-circuit current, and to apply the most stringent discrimination tests to the circuit-breakers. The maximum short-circuit current measured during the trials was 33 000 amp with the four 300 kW generators in parallel, and this was just about the maximum which the type of ring-main switchgear then in use could interrupt safely. Discrimination between the circuit-breakers was found to be unsatisfactory, and, to correct this, new over-current relays had to be designed which could be fitted in existing ships. High-breaking-capacity fuses were also fitted as back-up protection in branch breakers, and switchgear manufacturers were given the task of developing breakers of higher breaking capacity for future installations.

The urgent problem was how to deal with the new battleships which were then being built with eight generators, which, when run in parallel, would produce a short-circuit current beyond the breaking capacity of the switchgear in use. It was decided to run the ring-main system split in these ships so that no more than four 300 kW generators would supply one side of the ring main. One of the original objectives of the ring-main system, that by running with a closed system only limited reserve generator capacity was necessary, had therefore to be sacrificed, and subsequent systems were considered as a number of sections, each carrying its own reserve generator capacity, which could be interconnected in an emergency if any generators failed.

During this period the growth of demand for power was steady but not spectacular. Early hesitation over the use of electric motors for auxiliary drives was gradually overcome in all fields except for gunnery applications. Experiments with gun-driving motors were carried out during the first decade of the century, and in 1906 two 12-in gun mountings were installed in H.M.S. *Invincible* with electric training and elevating using the Ward Leonard system. The training motors were only 25 h.p., and it is not surprising that after a short time in service the installation proved to be under-powered. Attempts at electric drive of gun turrets were abandoned until just before the 1939 War, when the metadyne system was introduced to solve the problem of remote power control. To this I shall refer later.

Not long after the very early days when crude batteries were used to provide current for firing guns, there began a steady growth in the use of electrical equipment and instruments for passing gunnery data and orders from range-finder directors to guns, and for communication generally. The increasing work involved in dealing with this side of electrical engineering was greatly accelerated during the 1914–18 War. Many of the components of fire-control systems such as plotting tables and data-transmission and communication instruments were the products of specialist firms, and this made uniformity of practice and design difficult, if not impossible. In the years between the two great wars this shortcoming was put right, and more efficient and reliable systems and their components were developed and designed in detail within the Admiralty. Much of this progress was due to the genius and leadership of Mr. H. Clausen.

One result of this effort was the installation of Admiralty-designed fire-control computers in H.M. Ships *Nelson* and *Rodney*. A standard fire-control computer and fuse-setting unit were also designed for installation in small ships. As the air menace at sea developed we thus had a background with which to tackle the new and more intricate problems of designing fire-control systems and predictors for high-angle anti-aircraft guns.

Many electrical systems and devices were designed and used for improving the control of armament, and it is difficult to single out any one for special mention. One innovation of interest was Henderson's gun-director system, which was introduced towards the end of the 1914–18 War. This made use of a gyroscope to ensure that the guns fired only when the ship was in the horizontal position when rolling. This was done by stabilizing the firing contacts and was a prelude to the day when director and gun mountings would themselves be stabilized, thus permitting a very much higher rate of fire, independent of the movement of the ship.

Much work of far-reaching importance was also undertaken during the period under review at the Admiralty Research Laboratory. Three items are of particular interest:

(i) The plotting table, which automatically records the position of a ship on a chart.
(ii) The magslip for data transmission and reception.
(iii) An electrically-controlled oil-operated remote power system for searchlights.

The magslip is a precision selsyn type of data-transmission unit and was capable of very much greater accuracy than any other instrument available at that time and for many years later. It is now being superseded by the synchro. Detailed manufacturing drawings were prepared by the Admiralty, thus enabling production to be put in hand with a minimum of delay to meet expanding requirements in war-time. It was adopted by the Army and R.A.F., and production during the war was at a rate of many thousands per week. It set a new standard of reliability and accuracy in the transmission of signals in gunnery systems, and was later to be of particular value in providing the initiating and controlling signal for the remote power control of weapon systems to which reference will be made later.

The hydraulic control unit was adopted for the servo control in naval gunnery systems and by the Army for remote power control from the predictor of Bofors anti-aircraft guns.

1939 to 1945

In planning for war it was expected that the immediate emphasis would be on production of ships and of electrical equipment for new construction, refits and repairs. It was also expected that large-scale air attacks on London would disrupt the work of the highly experienced design staff working at the Admiralty. Plans were accordingly made for the immediate dispersal of the design teams to assist on production. This was done, but hardly had they arrived at their destinations when they were recalled to Headquarters, which had in the meantime been moved from London, as it very quickly became apparent that extensive effort would have to be devoted forthwith to research, development and design to meet and overcome the enemy's novel and ingenious devices in the war at sea.

The enemy began laying magnetic mines during the first week of the war, and ships were lost and damaged whilst counter-measures to protect ships and destroy these mines were being perfected and the necessary equipment produced. Although sweeps of a sort were produced at once, really effective sweeps could not be evolved until the characteristics of the German mine were revealed after a live mine was recovered and examined in November, 1939.

The towed skid and the magnet ship were the two most successful of the early sweeping devices. The skid consisted of a raft on which was mounted a coil energized from a wooden towing vessel at a safe distance from the skid. The magnet ship, or 'Borde', as this type of vessel was called after the first ship to be so equipped, had an electromagnet built into its bow. The magnet was 100 ft long, weighed 500 tons, and was supplied from a d.c. generator. Later the supply was changed to low-frequency alternating current so that the sweep could deal with mines of north or south polarity. The use of these ships had to be discontinued when the mines came to be armed with a device requiring an unknown number of actuations before exploding. Real success came when the towed buoyant-cable sweep was developed. The method consists in producing a large magnetic field in the sea by pulsing currents of several thousand amperes through a loop of buoyant cable towed by the minesweeper. It is still the standard method of sweeping magnetic mines.

Two types of buoyant cable were designed literally overnight by two firms whose names are (or used to be!) household words in the cable industry, and in each case the inventor was a Past-President of this Institution. So successful were these designs that, with relatively minor improvements, they continued to be used until the end of the war.

In addition to the problem of sweeping magnetic mines there was also that of making all ships safe by fitting degaussing coils. To meet the immediate and urgent requirements at the beginning of the war these coils were of a temporary nature and consisted of turns of cable sewn up in canvas and clipped outside the hull. Current was obtained from the ship's main electrical supply and adjusted to give the correct magnetic effect by resistance control.

The first permanent solution used a single-turn coil of rubber-insulated heavy copper strip fitted to the external plating of the hull. This, however, proved to be very vulnerable to damage even in moderate seas. After considerable research in 1940 it was decided that the degaussing cable could be effective when fitted inside the hull in spite of the magnetic screening which it had been feared would be excessive. This improvement, using conventional cable, made it possible to work at a higher voltage, and the modern system, using multicore cables with the conductors connected in series and supplied from the ship's main electrical supply, came into use with automatic adjustment of current with change of the ship's course and latitude.

Right up to the end of the war there was no relaxation in the mine and mine counter-measure contest. The enemy developed more sensitive magnetic mines with anti-sweeping devices, and later an acoustic and a pressure mine. These were the most difficult mines to sweep and destroy.

One of the first warships to suffer from a magnetic mine explosion was H.M.S. *Belfast*, in October, 1939. The hull structure was severely damaged as a direct result of the explosion, but apart from this the extensive damage to the electrical equipment from shock effect was sufficient to immobilize the ship. The first reaction was one of dismay that so many items were incapable of withstanding the shock effect of an underwater non-contact explosion. It was clear that the mechanical shock experienced in a ship as a result of such an explosion was far more severe than had been anticipated. Immediate action was taken to correct these deficiencies in design.

Designing electrical equipment to withstand shock was not new. It had been the practice for many years to do so, and to test prototype equipment on a shock-testing machine; but the impact allowed for was based upon the shock resulting from gunfire and the detonation of other weapons *in* the ship. Before this testing machine was designed, indeed as far back as 1913, important items of electrical equipment, such as circuit-breakers, had been tested for shock by mounting on a 5-in armour plate and firing a projectile at the plate from a 6-in gun at a range of 1 900 yd.

However, the shock accelerations and displacements experienced as a result of non-contact underwater explosions were of a much higher order, and work was put in hand immediately to design a new shock-testing machine to the new criteria. Fortunately, the *Belfast* provided valuable data because it was possible to pick out certain identical items of equipment, dispersed through the ship, some of which had suffered damage and others not. Tests producing corresponding damage enabled the magnitude of the shock to be established empirically. We were then able to modify and improve the electrical equipment itself or arrange to erect it on suitable shock-proof mountings. Very extensive use was made of a new testing machine which was designed and manufactured at the Admiralty Engineering Laboratory and is shown in Fig. 6. Later, similar machines were

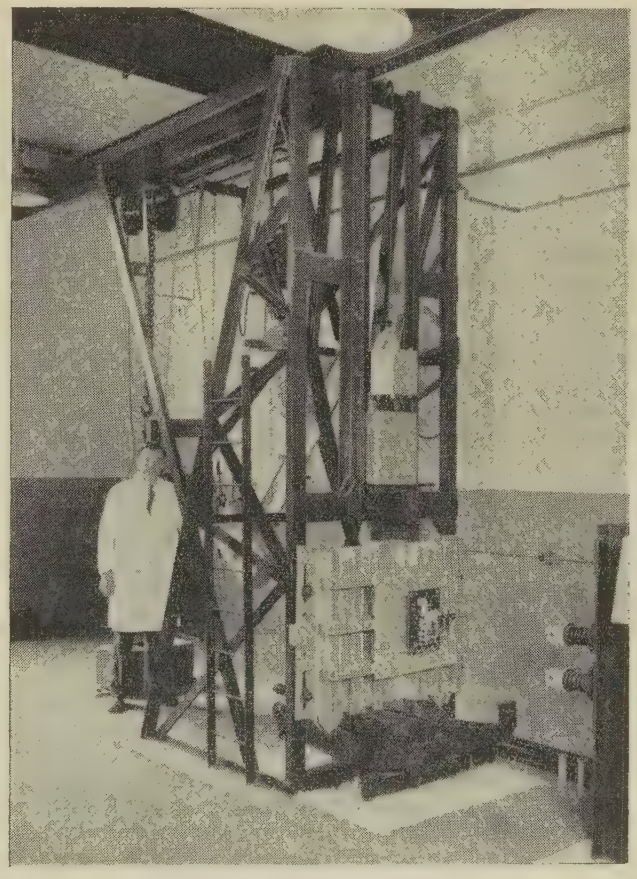


Fig. 6.—Standard mechanical-shock machine.

installed in the works of main manufacturers, and drawings were passed to the United States Navy so that they might benefit from our experience.

The evaluation of shock in ships is now an important and continuing part of the work of the Ship Department, and shock assessments are made in new classes of ships by fully instru-

menting the ships and subjecting them to underwater explosions of known intensity.

It was a great blow when H.M.S. *Ark Royal* was sunk in November, 1941. Although hit by one torpedo only, the result of the explosion was disastrous because it resulted in the immediate flooding of a boiler room and the main switchboard room. The flooding of the main switchboard, which was the control centre for the ring main and all outgoing services therefrom, was the vital factor, because, although steam was maintained for some time in the undamaged boiler room, it was impossible to close the circuit-breakers by remote control from the main switchboard to restore services. Hand operation of the breakers was impossible because of the state of the ship.

This situation had arisen because the control of the breakers from the main switchboard was based on a system in which the control wiring was always energized, and flooding or short-circuiting of this wiring could therefore change the state of a breaker by simulating the operation of either an 'on' or an 'off' push.

It therefore became of paramount importance to develop a new system. The outcome was a relay system in which the energizing power to operate or maintain the breakers in their open or closed state was contained within the breaker enclosure. Any change of state was achieved by energizing the control coil of a relay, which otherwise was dead, and which received its supplies from an independent low-voltage source at the controlling position, and in emergency from a secondary battery. The new system also made it practicable to introduce secondary control positions for sections of the ring main distributed throughout the ship, which could be brought into operation immediately if the main controlling position became unserviceable. Fig. 7 shows a typical main switchboard which contains these features, and Fig. 8 shows one of the four secondary control switchboards.

The loss of the *Ark Royal* emphasized another very important factor in the electrical supply for warships, namely the necessity for having generators capable of maintaining essential supplies after loss of steam.

The fact that the steam turbine was the most reliable prime mover was the main reason for a Board decision in 1938 to fit steam-driven generators only in future ships. The experience in the *Ark Royal*, however, demonstrated the need for alternative prime movers, and a programme of installing Diesel-generators in those ships where none existed was immediately put in hand.

The extent to which reserve electrical power should be provided in a warship had been the subject of much debate over a number of years, but it had now become clear that sufficient reserve must be available, over and above that required to meet the normal action load, to maintain essential and important action and salvage services in the event of damage. Automatic change-over switches to switch on alternative power supplies were introduced for close-range anti-aircraft gun mountings to ensure the minimum interruption of electrical supplies, and hence of gunfire, in a damaged ship whilst under attack, and their use was soon extended to other important services. The largest of these switches was able to transfer automatically a load of 110 kW in a quarter of a second on failure of normal supply. The use of these automatic switches, and indeed the hand-operated change-over switches which were also used, demanded that reserve electrical power should be available to meet the resulting transferred loads. Here again was justification for standby

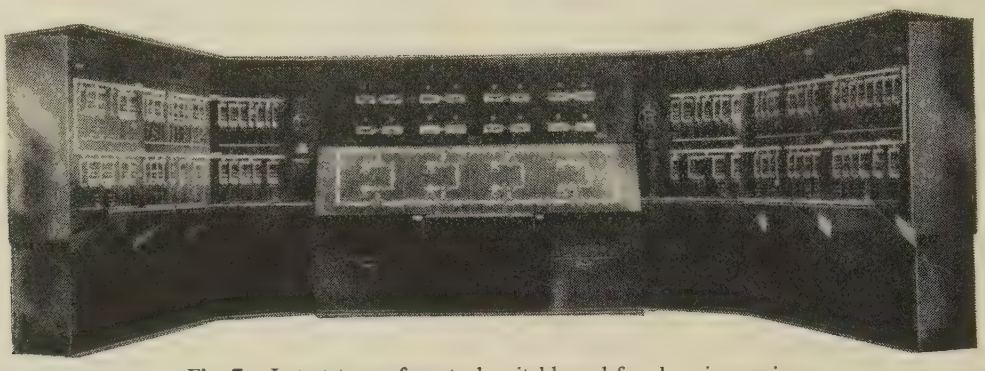


Fig. 7.—Latest type of control switchboard for d.c. ring main.

Diesel-generator capacity to ensure that a warship could maintain essential services, including maximum gunfire power, even though steam were lost. Throughout the war and the immediate post-war period, additional generators, both steam and Diesel, were installed in serving ships to keep pace with the increase in power demand and to ensure adequate reserve.

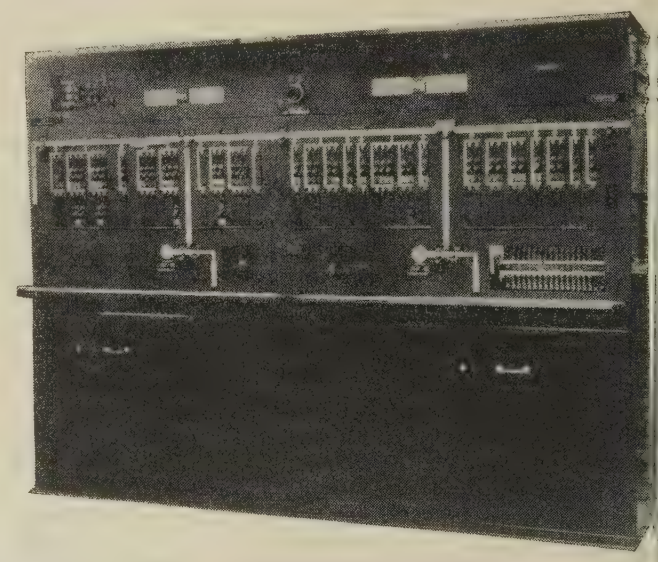


Fig. 8.—Secondary control switchboard.

The part played by electrical engineering at sea during the war was much greater than earlier experience had led us to expect. The development of remote power control systems for automatic control of armament in the early months of the war is a good example, electrical systems for this purpose having lain dormant since 1906. The objective was to eliminate the inaccuracies and time lag inevitable in existing systems in which processed data were made available to the gunlayers by mags pointers giving angles of elevation and training. It was the

he task of the operators so to operate the gun if manhandled, r in the case of heavier gun mountings the hydraulic local ontrol, that magslip pointers indicating elevation and training vere kept in alignment with the data-transmitting pointers. Such system was acceptable for surface gunnery but was quite nadequate for dealing with rapidly moving aircraft targets. Vork was put in hand in 1939 to adapt the metadyne system hen in use for traction to the direct control of gun mountings. The first application was to multi-barrelled pom-pom mountings or close-range anti-aircraft defence. The system was based on nagslip transmission and was progressively applied, following s success with the pom-pom mounting, to various calibres of un mountings up to and including 6 in, and also to the director owers which control the guns. Briefly, a magslip data system roduces an alternating-voltage signal which is proportional to ne misalignment between the transmitter on the gun director nd the receiver on the gun mounting. This signal is fed into n amplifier where it is rectified in such a manner that the olarity of the resulting d.c. signal depends on the direction of nisalignment, phase-advanced to provide stability, and finally mplified, and applied to the field of the metadyne generator as current signal. The metadyne gives further power amplifica- on, and the resulting output is applied to the armature of the eparately-excited constant-field gun-driving motor. Training nd elevation are, of course, treated separately.

The 'brains' of the metadyne system, namely the amplifier, rocessing error-in-alignment signals, was later successfully dopted to control the valves of hydraulic-powered mountings. een competition was maintained between the metadyne and its ydraulic rival and both are in use to-day.

When radar was introduced for range finding it was an obvious leal to apply automatic following of the radar aerial to the netadyne-controlled director, which then became a radar aerial arrier. So the radar aerial, the gun director, and hence the un, automatically follow the target. Later the radar was esigned so that it provided its own misalignment signal and ne aerial automatically 'locked on' to the target. Thus fully utomatic following of the target by the gun mounting was chieved and all the necessary additional information for ccurate gunnery was fed into the control systems from the abilizer and the fire-control computer.

In the foregoing, reference has been made specifically to the etadyne as the cross-field generator used by the Royal Navy. he metadyne's twin brother, the amplidyne, is also in use for arious applications and in particular for radar aerial positioning.

These developments inevitably led to an increased demand for ectrical power because the requirements of stabilization and mote power control led to higher speeds and greater accelera- ons in following fast targets, while at the same time the number f close-range anti-aircraft guns in ships increased to meet the ver-growing menace of air attack.

The control of a warship under action conditions, in particular fter damage, calls for, and indeed the safety of the ship may epend on, an efficient damage control organization. The rganization now in force in the Navy was evolved, and proved s worth, during the war. Of course such an organization tends far beyond the control of electrical power. It includes ne mechanical and hull departments of the ship as well, but its ffectiveness depends upon good and reliable internal communi- ation, so here again the electrical engineer is primarily involved.

The absence of reliable secondary lighting in the early days f the war was responsible for the loss of many lives. Oil nergency lanterns proved completely useless under damage onditions and were replaced by battery-supplied lanterns, con- olled by a hold-off relay energized by the ship's main supply, hich come into operation automatically. When the *Ark Royal* was sunk she had 35 of these automatic lanterns, the most that could be made available at that time. The present *Ark Royal* has 800 lanterns carefully sited throughout the ship to ensure that every passageway and compartment is sufficiently illuminated for safety should the main supply fail.

Another war-time innovation was the action information organization requiring an action information centre. This compartment has become the most concentrated part of the electrical installation in a warship. It is the focal point for radar and asdic information, external and internal communication, aircraft and weapon direction, and navigation. Furthermore, the growth of the radio installation, as distinct from radar, was phenomenal, and an increase in the number of transmitting and receiving channels, together with remote operation of radio-communication equipment situated in various parts of a ship, brought in its train still further increases in power demands and in the number of cables for these systems.

The increasing dependence on data systems, and on the control of electric power generally, emphasized the need for careful installation in a warship to ensure the maximum possible protection against action damage. The warship had now become entirely dependent for its seagoing and fighting efficiency upon its electrical installation, and in recognition of this fact every effort was and is made to site cables to take advantage of the greatest possible protection and to duplicate supplies by alternative routes.

It was anticipated that a large amount of damage sustained by our ships in the event of war would be by aircraft bombs, but the pattern of damage from bomb attack proved to be somewhat of a surprise. The electrical layout had hitherto been based on knowledge of the effect of shell bursts, and alternative supplies from opposite sides of the ring-main system in a ship of reasonable beam had been regarded as adequate. War experience, however, revealed that the splinter pattern following a bomb explosion differed from that following a shell explosion, being no longer almost entirely directed forward owing to the velocity and trajectory of the shell when it penetrates a ship. Electrical fittings and cables on the side of the ship presented to attack usually enjoy a reasonable degree of immunity from damage by shell splinters; so that, if important cables are arranged as separate and distinct systems on each side of the ship, one system has a reasonable chance of survival. Bombs which penetrate a ship have a relatively lower residual velocity when they burst, and this results in a substantially horizontal belt of splinters which, on the deck level at which the bomb bursts, may do damage right across the ship unless the beam is very large. The circuit-breaker rooms on opposite sides of the ship had therefore to be staggered in order to provide a degree of longitudinal separation of sources of alternative power in addition to separation athwartship, and some vertical separation of alternative runs of important cables was also found to be necessary.

This new pattern of splinter damage, together with the experience of flooding and extensive areas of damage resulting from underwater torpedo explosions, raised considerable doubts as to the value of the ring-main system in warships. These factors were kept very much in mind when the change to an a.c. system after the war provided an opportunity for a new approach to the supply and distribution problem.

Nevertheless, tribute must be paid to the ring-main system which was introduced 50 years ago, and which is still in existence in some of our larger ships. Its great merit was economy in cables and bulkhead glands, because distribution was in general vertical from the ring main and conformed with the watertight subdivisions of the ship.

Many other lessons were learned as the war progressed, and

improvements were continually being made to the electrical systems and equipment throughout the whole course of the war at sea. Many years have passed since then, and during the years of peace it has been necessary to exercise unceasing vigilance to ensure that the lessons learned in those days are not forgotten, as well as to look into the future to try to anticipate problems which may arise in the ever-changing pattern of war at sea. Many aspects of design are not compatible with each other; the most diverse features are functional efficiency as warships and habitability to meet the creature comforts of those who man them.

Post-War to the Present Day

The ever-growing demand for electrical power made us realize that the ring-main system, as we knew it, could not cope with any substantial increase in generating capacity. The limit of capacity of individual generators with existing ring-main switchgear was 500 kW, and we now needed to think in terms of 1 000 kW generators at least for future cruisers, aircraft carriers —and indeed even for smaller ships—if we were to keep the number of generators in the system down to acceptable limits. To deal with these larger powers we had to consider the use of a voltage higher than 220 and whether it should be direct or alternating.

The number of motors in ships kept increasing (there are over 1 000 in a large carrier), and the replacement of the d.c. motor, with its commutator and brushgear, by the rugged squirrel-cage motor became an obvious and worth-while objective from the point of view of reliability and maintenance.

The position was now different from that in 1932, when a change to alternating current had last been considered and turned down, and with our war-time experience and the larger electrical installations we were able in 1946 to reconsider the question on a much broader basis. Furthermore, the ever-increasing demands to accommodate more and more operational equipment in warships made saving in weight and space of paramount importance.

The investigation showed that weight could be saved, even in a relatively small warship with a total of 1 000 kW of installed generator capacity, by adopting a 440-volt 60 c/s 3-phase a.c. system, with a proportionately greater saving in weight as the size of the electrical installation increased. Such a system had been in use in the U.S. Navy for some time. Although the possibility was considered, a voltage higher than 440 was not selected. With 440 volts for the basic system a single-phase voltage of 115 was adopted for lighting and small power loads. This voltage was chosen because the lamp filaments would be more resistant to mechanical shock and vibration, the design of miscellaneous small power-consuming devices was made easier, and there would be rather less danger from electric shock from small equipment supplied at that lower voltage.

The choice of frequency was not easy. Theoretically there is a case for adopting a frequency higher than 60 c/s in order to obtain further reduction in the sizes of motors and transformers, but the conclusion was reached that a wholesale departure from available standard designs would be very difficult to justify.

Investigations showed that a motor speed of 1 800 r.p.m. would meet a large number of requirements, and this, together with the advantage of 60 c/s in reducing the sizes of generators and transformers, made that frequency very attractive. Naturally, thought was given to adopting the widespread standard frequency of 50 c/s, particularly to make it easy to supply electrical power from ashore; but the lower speeds would have involved a quite unacceptable weight penalty. So, with the full support of industry when the facts were put before it, 60 c/s was selected. The added advantage of standardization with the U.S. Navy is obvious, and since we made our decision the other Nato navies have followed suit.

Although there is no alternative in cruisers and larger ships because the size of the electrical installation demands a distribution voltage higher than 220 volts, and therefore preferably an a.c. system, the case for alternating current in the smaller ships rested on:

(*a*) Reduction in weight.
(*b*) Considerable reduction in maintenance effort—an important feature in a small ship in which maintenance facilities are limited.
(*c*) Uniformity of design and operating practice with bigger ships.

An a.c. system is, of course, handicapped by its inability to provide, in a simple way, the speed control of motors which is readily available with direct current. This difficulty has been overcome to some extent by persuading the designer of the driven machinery that several speeds are not essential or, at the most, that a choice of two—obtained by pole changing—would meet his needs. For certain equipment, of course, a range of control of speed is essential and the extra weight and complexity have to be accepted.

The investigations which led to the adoption of alternating current in warships of destroyer size and above were made in connection with the design of a new class of cruiser and the *Daring* class destroyers. No orders were placed for that class of cruiser, but eight *Daring* class destroyers were ordered, and approval was obtained to build four with d.c. systems and four with a.c. systems so that a direct comparison could be made in ships where the advantages of a change to alternating current were marginal. It was intended to await operating experience with these ships before finally considering a wholesale change to alternating current; but before such experience was available it became necessary to design systems for three new types of frigates, and it was decided in 1948 to adopt alternating current for these and all future ships larger than minesweepers. Special considerations apply to minesweepers which favour the continued use of direct current.

The first large ships to have an a.c. system were the *Tiger* class cruisers, the construction of which had been suspended at the end of hostilities, because when it was eventually decided to complete these ships with modern armament and radar the consequent large increase in demand for electrical power had to be met by substituting four 1 000 kW a.c. generators for four 500 kW d.c. generators.

The recently completed aircraft carrier H.M.S. *Hermes* was dealt with in a similar way, and Fig. 9 shows the switchboard system installed in that ship. Within the limits imposed by a design not originally intended to accommodate a large a.c. system, the generators and their switchboards are grouped as unit close to their associated boiler rooms.

When working out the details of the new a.c. systems it was necessary to lay down voltage regulation standards. It was decided to adopt a system in which the generator voltage is controlled to fine limits by using rapid-response automatic voltage regulators in association with appropriate exciters, and a cable system designed to give the minimum practicable voltage drop. Thus designers of individual items of equipment would have the benefit of a closely controlled supply. The alternative which was rejected was to control the generator voltage by more conventional methods giving a wider range of voltage regulation and at the same time to allow a wider range of voltage drop in cables. This would have meant that, where necessary, the designers of certain equipment would have had to incorporate voltage-stabilizing equipment in their apparatus to provide the closer voltage limits.

It was decided that the voltage at the generator terminals should be maintained within the limits of $\pm 1\%$, and that the

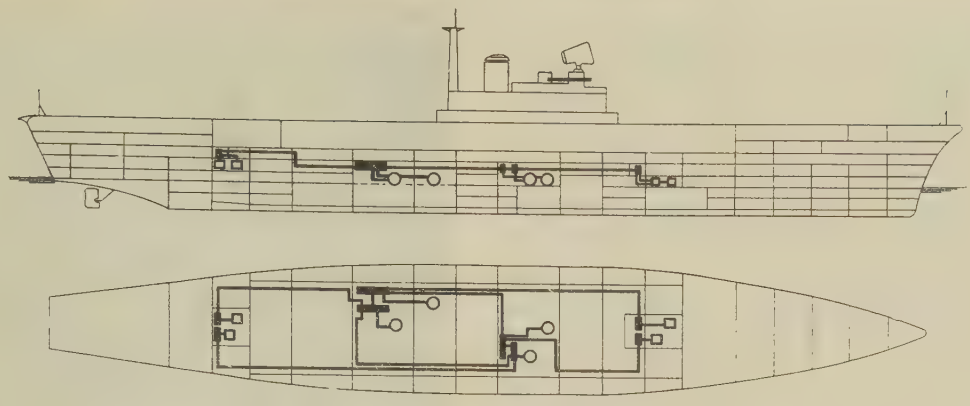


Fig. 9.—A.C. main supply system in H.M.S. *Hermes*.
440 volts, 3 phase, 60 c/s.
■ Switchboards.
□ 360 kW Diesel-generator.
○ 1 000 kW Turbo-generator.

argest voltage drop allowed on the system due to starting nduction motors should not exceed 15%. Furthermore, after his voltage drop, recovery to ±1% was designed to be achieved within 0·3 sec. Later, it was found that a recovery time of 0·5 sec was acceptable. Frequency standards were also laid down.

Some converted supplies at higher frequencies are inevitable n a warship, but every effort is made to keep the amount of conversion equipment to a minimum. The higher frequencies selected are 400 c/s and 2 400 c/s. These are nominal and are derived almost entirely from generators driven by induction motors which produce a near multiple of 60 c/s. Direct current s still required for such services as degaussing, field excitation for special motors, computing and telephones. These supplies were largely provided by motor-generators in the first a.c. ships. Selenium rectifiers were also used and are still in use. They in turn are being superseded by silicon rectifiers, where appropriate, with a significant saving in weight and wild heat.

The 440-volt 3-phase system is a 3-wire insulated system with no earthed neutral. This differs from shore practice, but the essential feature in a warship power system is to maintain supplies in an emergency, and it is thus necessary to avoid tripping of circuit-breakers as a result of an earth fault on one line. Earth faults are indicated on the switchboard and are dealt with by the maintenance staff as soon as possible. The aspect of shock to personnel was not lost sight of when the all-insulated system was adopted. With alternating current a voltage over 50 can be fatal, and although by introducing an earthed neutral the possible voltage to earth is reduced, even this reduced voltage of 260 volts is highly dangerous in a ship. All 440-volt equipment is therefore designed to prevent accidental contact with live metal and is provided with 'difficult access' covers. The main switchboards, of which that shown in Fig. 10 is typical, are totally enclosed with dead fronts, and the breaker and switch-fuse units are withdrawable for servicing. It is very reassuring that so far we have had a casualty-free record.

Portable equipment is supplied from 115-volt single-phase systems which are unearthed. In order to increase the safety of such equipment, the circuits are taken from small transformers (not greater than 5 kVA) reserved for such use, or from reserved separate phases of larger transformers which have open-delta secondary windings. The ship's lighting and miscellaneous small-power equipment have their own transformers, so that the possibility of earth faults on the portable equipment socket networks is reduced to a minimum. A 50-volt supply from special transformers is provided for boiler-cleaning equipment because of the enhanced risk of shock when performing this work.

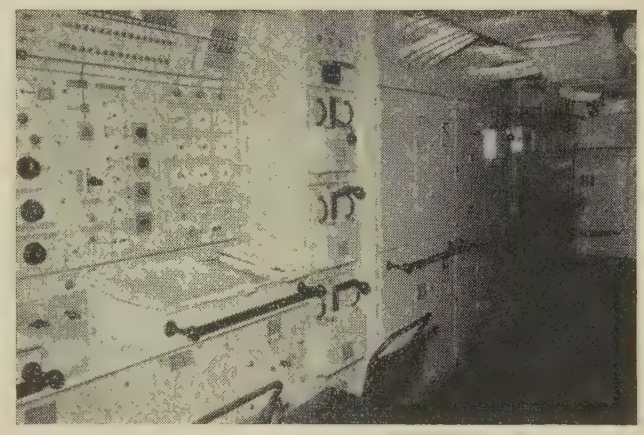


Fig. 10.—Typical a.c. switchboard.

It may be of interest to remark on a few of the special design problems which arose and which continue to demand our attention as a result of the use of alternating current.

The design of the circuit-breakers and motor control-gear for the first a.c. ships called for the closest collaboration between the Admiralty and industry because of the difficulty in making equipment of this type proof against shock from underwater explosions. For motor control-gear the problem could have been eased by using d.c. operation from rectifiers; but the aim was to save weight and space wherever possible and to provide simple, easily maintained equipment. Few firms were able to meet requirements when the first a.c. ships were being built, but that position has now been rectified and we can call on many more firms for this specialized equipment.

Single-speed squirrel-cage motors with direct-on-line starters are used wherever possible. In the first ships a few reduced-voltage starters were necessary, but the later ships have larger generators and direct-on-line starters can be used in all instances. A few of the larger motors are designed for low starting currents, thus avoiding the necessity for complicating and adding weight to the control gear with voltage-reducing devices.

The method of obtaining over-current discrimination with the

d.c. breakers in the ring-main systems was a well-tried one, and the same method was adopted for the hand-operated electrically-retained, and the electrically-operated, circuit-breakers in all a.c. ships to date, by using d.c. solenoids supplied from rectifiers. Short-circuit trials carried out in H.M.S. *Diamond*, one of the *Daring* class destroyers, showed that the system was not ideal for a.c. ships, and improvements have been incorporated in new designs of circuit-breakers for later ships to ensure more reliable discrimination and better protection of prime movers against severe overloads.

The introduction of alternating current brought with it the problem of voltage regulation of generators. With the relatively stable d.c. generator, pushbutton control of voltage by the switchboard attendant was acceptable, and in the d.c. era radar and other equipment requiring close-voltage a.c. supplies were catered for by motor-generators with carbon-pile voltage regulators. With a.c. generators some degree of automatic voltage control was accepted as being essential, and, that being so, it was decided to adopt a system of automatic voltage regulation which would meet the requirements of all basic-frequency equipment in the ship. The first a.c. ships used a thermionic voltage regulator which was duplicated, with means of automatic change-over from one channel to the other. The output was used to energize the control winding of a cross-field exciter. This system gave a voltage regulation of better than 1% with a recovery period of 12 cycles following the throwing-on of the test load of half full-load current at a low power factor.

The accelerated building programme in the early 1950's resulted in several generators of different outputs and several new designs of voltage regulators having to be put into service before experience with the earlier type of regulators had been gained. One of the designs incorporated a magnetic amplifier and this has given good service. In 1957 representatives from industry and from the Admiralty collaborated in the development of a standard design of automatic voltage regulator employing a single-channel magnetic amplifier. Power for the amplifiers is taken from an auxiliary 1 600 c/s permanent-magnet generator, which ensures continuous operation during the recovery period following a severe short-circuit. These regulators can be used with single exciters of conventional design to cover a range of generators from 150 kW to 1 500 kW.

Warship electrical systems are designed with a view to maintaining electrical supplies even though damage occurs in action. In spite of this, war experience proved the value of a system of emergency cables and connectors which can be used when damage to the permanent wiring is extensive. The problem of designing an emergency cable system for use at 440 volts a.c. under the conditions likely to prevail after severe action damage was no easy one. It has been solved by using plug-in connectors on the ends of the flexible cables, the connectors being fitted with safety sleeves which retract when the plug is inserted in the appropriate socket. The plugs can be inserted only in the correct phase socket, and are locked by an insulated key. Provision is made for emergency supplies to be taken from the switchboards by means of a special type of socket on the switchboard, and arrangements are made so that important equipment can be given an emergency supply through sockets adjacent to the motor starters or in the fuse panels supplying groups of equipment. It is necessary, of course, to provide special watertight bulkhead terminals so that the emergency power can be taken through the ship without having to leave watertight doors open. Coils of prepared cable, fitted with the protected plugs, as shown in Fig. 11, are dispersed throughout the ship. Portable fuse panels and transformers for lighting are also provided.

There is no doubt that the decision to adopt alternating current

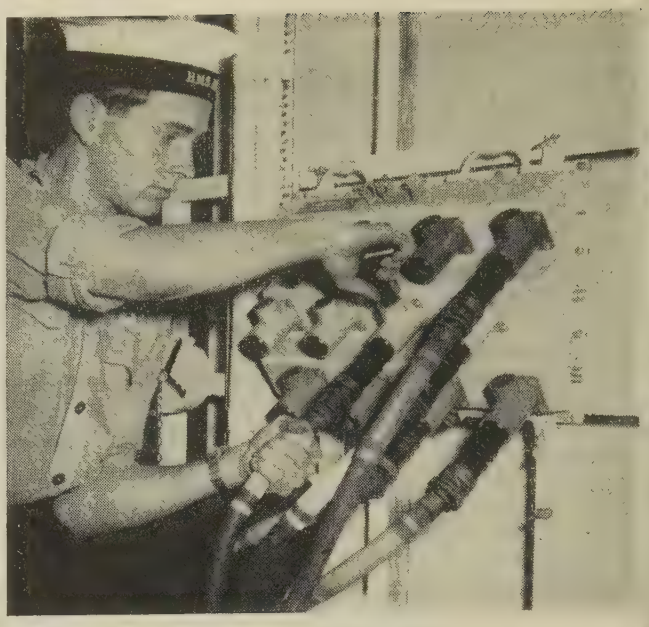


Fig. 11.—Emergency cables.

was right. Admittedly, many problems have arisen and not all have been completely solved. Motor reliability is not yet all it should be, but new insulating materials will help us towards our objective of greater reliability.

Before concluding this review of post-war progress, it seems appropriate to make some reference to the selection of prime movers for driving generators. In this aspect of warship design, as in many others, the requirements of peace and war are conflicting. In peace it is necessary to reduce watch-keeping in harbour to a minimum so that it is desirable to let steam die down and change over from steam turbo-generators to Diesel generators. Thus the peace-time role of the Diesel-generator is to maintain supply in harbour.

In war it is normally necessary to keep up steam even in harbour, which means that the steam turbo-generators can continue in use. The function of the Diesel-generators then becomes that of providing a standby or auxiliary source capable of providing sufficient power, in the event of all steam being lost, to supply the salvage load, with enough left over for at least the anti-aircraft armament so that the crippled ship may still defend herself. These separate functions are not always compatible, and as in other aspects of warship design the best compromise has to be reached. The gas turbine is capable of fulfilling a dual role, and with that in view some of the generators in our latest ships—the guided-missile destroyers and the general purpose frigate—are driven by gas turbines.

The demand for electrical power in warships has increased by leaps and bounds. Fig. 12 shows the growth of generating capacity in various types of ships over a number of years. Part of the increase in later years is admittedly due to a greater acceptance by all concerned of the need for more reserve than was usual before the war. Apart from the obvious growth in the use of electricity for motor-driven auxiliaries, particularly since the advent of alternating current, there has been a steady increase in the operational power demanded for weapon systems, radar and radiocommunication. The domestic electrical load continues to grow, ranging from air conditioning to drinking water coolers. The modern sailor expects reasonable amenities and to be able to use his own radio set, record player and electric shaver. In the general drive to improve amenities in warships, electricity plays a big part.

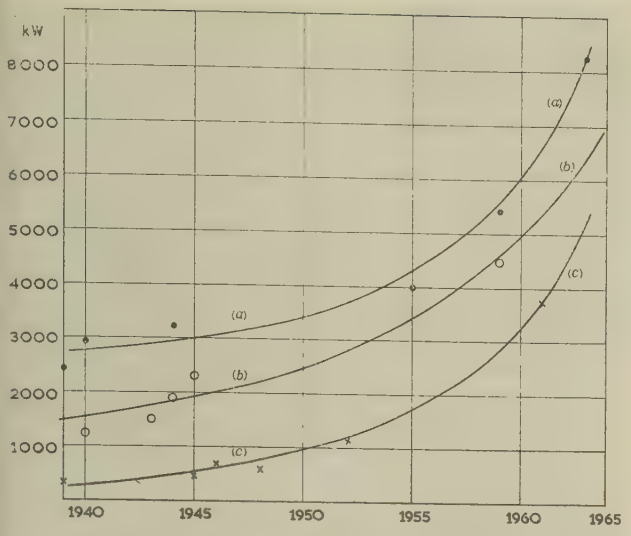


Fig. 12.—Growth of generator capacity.
(*a*) Aircraft carriers.
(*b*) Cruisers.
(*c*) Destroyers.

Fig. 13A.—Petty officers' accommodation in H.M.S. *Hermes*.

Fig. 13B.—Petty officers' accommodation in an older ship.

The electrical equipment in a modern warship encroaches on all compartments, but modern practice demands concealment as far as possible in living spaces. Fig. 13A shows petty officers' accommodation in the *Hermes*, where electrical installation is not so much in evidence as in earlier ships, of which Fig. 13B is a good example.

In aircraft carriers there is a heavy demand for electrical power at various voltages and frequencies in the aircraft hangar and on the flight deck for servicing aircraft. Also, the electrical equipment associated with airborne radar and missiles requires to be energized for long periods whilst awaiting take-off to avoid unnecessary running of the aircraft's engines.

In this post-war section I must make some reference to other improvements and progress in various aspects of naval electrical engineering. A much higher standard of demagnetization is now necessary in minesweepers to ensure safety when sweeping in shallow waters. Even the eddy currents induced in the metal frames of composite wood and aluminium hulls when rolling in the earth's magnetic field are capable of producing fields which can fire a sensitive magnetic mine. These unwanted eddy-current fields have to be compensated. Detector coils, fitted at the top of the mast, provide signals which, suitably amplified, initiate compensating currents in the ship's degaussing coils.

The development of the electrical analogue computing systems for modern anti-aircraft gunnery is an outstanding example of post-war naval electrical engineering. A distant aircraft target is detected and held in the sights of a director or its modern equivalent, a tracking radar set. The movements in elevation and training of the director or radar as it follows the moving target provide data from which the present position, speed and direction of the target can be computed. Solving further equations will predict its future position, and equating these to yet further predictions of where a shell will be at a given time after it is fired will yield, if the calculations are correct and completed in time, the necessary information as to where to point the gun so that the shell and the target will meet. It is also necessary, of course, to make accurate allowance for other factors, such as the rolling and pitching of the ship. A solution to this problem was reached by developing electrical analogue computers capable of continuous computation and prediction. This was a long and complicated task, but it was successfully accomplished and has resulted in gunnery accuracies hitherto unattainable.

To facilitate production and maintenance of such a system, it was divided into sections and further subdivided into a large number of individual plug-in units. Although a comparatively large number of amplifiers were required, it was found possible to confine them to a few basic types, which were arranged to conform to common dimensions, so that all would plug into a common rack.

Internal communications play a vital part in the running of a warship. They range from straightforward 500-line automatic telephone exchanges of the G.P.O. type used for administrative purposes, and installations for entertainment, to a series of operational communication systems. The most complex of all permits aural communication in an ambient noise level of 130 dB. The most obvious example of the need for this system is the flight deck of a carrier when the aircraft jet engines are running. The problem was solved by radiating signals from a loop round the flight deck. These are picked up by a miniature transistorized personal receiver, the output of which is fed to a telephone-type inset in the earpiece of a noise-excluding headset worn by all personnel concerned.

The scale of an internal communications installation in a modern aircraft carrier may be gauged from the fact that there are over 800 telephones and well over 600 loudspeakers. An

experimental closed-loop television installation has recently been completed for the briefing of air crews.

Since 1945 a major contribution to saving in weight has been made in cables by replacing the lead-alloy sheathing with a polychloroprene sheath—30 tons weight were saved in a destroyer of the *Daring* class. The removal of the metallic screening, however, raised doubts whether the interference with W/T reception would be unacceptably increased. Careful attention to suppression of interference at its source, and the use of 'in line' capacitors as filters, has avoided any increase in radio interference.

A more recent change resulting in a further saving in weight has been the introduction of silicone rubber as the insulant for all cables. This change was made possible by the co-operation of the British and American cable manufacturers after a detailed examination of American practice in this field. One of the advantages is that the residual ash of silicone rubber, after being subject to fire, continues to act as an insulant so that circuits continue to function even after compartments through which the cable are run have suffered damage by fire. The combination of polychloroprene sheathing and silicone-rubber insulant has made possible cables of reduced diameter and weight so that installation is greatly facilitated. The saving in weight and in the space occupied by cables is also valuable. The value of the change is well illustrated by Fig. 14.

Fig. 14.—Cable trunk.

A further recent improvement has been the introduction of a hanger system for the support of main runs of cables. This replaces the traditional carrier-plating system and has the advantage of simplifying installation, saving weight and providing easier access to the ship's side for painting. Fig. 15 illustrates the obsolete carrier-plate installation, and Fig. 16 the new cable hangers.

Fig. 15.—Carrier-plating system.

Fig. 16.—Cable-hanger system.

Consequent upon the ever-increasing numbers of cables which have to be installed in modern warships, a new form of watertight bulkhead gland has had to be developed to permit more cables to pass through the area available in bulkheads. This new gland takes the form of a box welded into the bulkhead through which the cables pass. The box is then sealed up solid by filling it with a quick-setting rubber-like compound. Such boxes are shown on the left-hand side of Fig. 17.

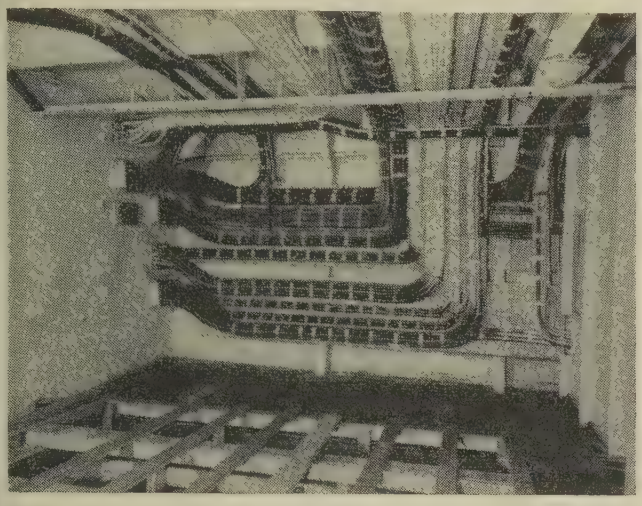


Fig. 17.—New method of passing cables through bulkhead.

It will be readily appreciated that the increase in the use of electricity for power purposes in ships results in a corresponding increase in the number of power cables to be installed. What is perhaps not so obvious is that there is an equal weight of cable used for operation and control purposes in the various weapon systems, including the associated radar, asdic and internal communications. The installation of the cable system for such services in H.M.S. *Eagle* involved making half a million connections. Fig. 18 shows the operations room in a frigate while cables were being installed.

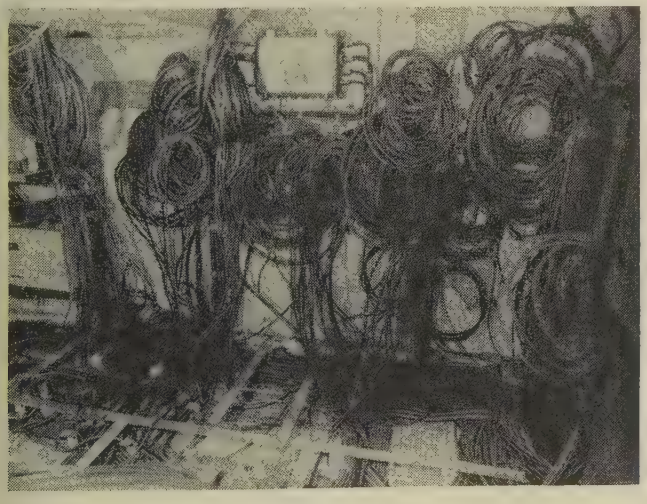


Fig. 18.—Operations room, showing the false floor.

I shall complete this post-war section with a brief survey of the special electrical problems associated with the submarine.

The war was fought with submarines in which, although there had been detailed improvements, the electrical system by which the vessels were propelled when submerged were substantially as they had been for twenty years. The main battery, in two or three sections each of 112 lead-acid cells, gave a nominal voltage of 220, and was only sufficient to permit maximum submerged speeds of 9 knots for periods of less than one hour. Speeds to permit reasonable submerged endurance had to be as low as 4 knots. The circulating air in the main motors was water cooled, and switchgear comprised open-type hand-operated knife switches protected by fuses. The limitation on submerged speed and endurance was a very serious handicap to submarine operations, and after the war great efforts were made to improve the underwater performance of new designs of submarines.

The first attempt made use of the Walther system, in which oxygen is stored in the form of highly concentrated hydrogen peroxide (high test peroxide) and used as the oxidant for burning oil when submerged. Two experimental submarines with this system are now in service, and until 1959 held the world record for underwater speed. The machinery, however, is complex and requires much maintenance. H.T.P. is a very dangerous and expensive liquid, and high speed endurance is limited to a few hours by the amount of H.T.P. which can be carried.

The second method of improving submerged performance is by increasing the capacity of the battery. War-time built submarines of the 'T' class were cut in two and lengthened. The increased space within the hull was used to accommodate a fourth battery section and two additional double-armature main motors, which were installed so that there are two motors (four armatures) on each of the two propeller shafts. For high speeds two battery sections are connected in series to give a nominal 440 volts. New cam-operated switchgear for the control of this equipment was developed.

The 'T' conversion, as it was called, was a stop-gap, and in the new *Porpoise* class a radical departure from previous design practice was made. There is one twin-armature main motor on each shaft, and these are used to propel the vessel on the surface as well as submerged. The whole power of the ship is controlled from one position, and this was an innovation in submarine electrical engineering practice.

For maximum speeds the batteries are connected all four in series to give 880 volts, and for ordinary working the system voltage is 440. Motor-driven cam-contactor switchgear is used, and special attention has had to be given to fault protection by means of high-speed high-breaking-capacity switchgear. To test the equipment, one of the largest-capacity d.c. short-circuit-testing installations in Europe was built at the Admiralty Engineering Laboratory. For transmission of power through the boats between batteries, main generators and main motors, copper busbars insulated with moulded micanite and metal-enclosed are used.

In electrically propelled submarines the battery determines the limits of performance, and by careful attention to design, thinner plates and improved separation, the capacity of the standard lead-acid cell has been increased from 4 750 Ah, at the 5-hour discharge rate, at the beginning of the war, to 6 560 Ah at the present time. Despite this increase in capacity an operational life of 300 cycles is still obtained.

There are, however, other problems for the battery designer besides specific ampere-hour capacity. Hydrogen gas is constantly being evolved from lead-acid batteries, and a mixture of 4% of hydrogen with air is highly explosive. It has been found possible to reduce the open-circuit evolution of hydrogen by the development of battery grid alloys with low antimony content. Research is continuing in an attempt to produce batteries completely free from antimony which will not evolve any hydrogen at all on open-circuit, and at the same time will eliminate the production of antimony trihydrite (stibine), a very toxic gas.

Water cooling has been adopted for the latest high-perfor-

mance batteries, and a further development is electrolyte agitation to permit shorter charging times.

Battery containers have been continuously improved to withstand shock resulting from depth charges. Ebonite containers were first superseded by laminated plywood with a rubber lining, and these in turn have been replaced by rubber-lined resin-bonded glass-fibre boxes.

The latest method of improving submarine performance is the introduction of nuclear propulsion plant which, by providing virtually unlimited power and endurance, will revolutionize submarine warfare. Indeed the submarine becomes at last a true submarine and not just a submersible vessel. Although there are many special problems, the electrical installation in a nuclear submarine in fact becomes more like that of a modern surface warship with the addition of an electric drive of small power for emergency use. Also, a nuclear submarine must still have a battery for starting up and emergency standby, although this is reduced to one section of 112 standard submarine cells. Because of the possibility of prolonged submerged operation the problem of hydrogen evolution assumes even greater importance than hitherto, and a further complication is introduced by the danger of accumulating stibine. The efforts to develop a reliable battery with antimony-free plates have therefore been increased.

Like surface ships the submarine has to accommodate an increasing amount of electronic and servo equipment for weapons, asdic and other purposes; and to this is being added, in the atomic submarine, new problems of reactor control and instrumentation. Space for all this equipment is difficult to find, but in the application of the semiconductor in its various forms lies hope for some easing of problems of space, reliability and performance.

The Future

I have attempted to review how electrical engineering has been applied to meet the requirements of the Royal Navy, and how it has developed over the last three-quarters of a century. I propose to conclude my review by very briefly attempting to forecast what further changes we may see, perhaps in my lifetime.

In the personnel field, I think it will be inevitable that the pattern of increasing specialization in design of equipment and installations must continue to an even greater degree. Both the serving naval electrical engineer and his civilian counterpart in the supporting services ashore will need to be of the same high standard or even higher than to-day. With the inescapable drive to reduce staff to the minimum, the need constantly to review organization methods to streamline and eliminate all unproductive work will continue.

In the material field, thanks to the close co-operation that exists between the Admiralty and the electrical industry, knowledge of new discoveries and developments becomes available at a very early stage so that possible naval applications can be kept in mind. Constant application of judgment will be increasingly necessary to decide when new materials and devices can be adopted for warship installations. The biggest drive in future years will continue to be for increased reliability and reduction in size and weight in every field of application. Already in the electronics field this has led to micro-miniaturization and the adoption of semiconductors for many applications. In the ever-broadening sphere of the semiconductor, possibilities range from microamperes to kiloamperes, embracing minute amplifiers and bulk power conversion. He would be a rash man indeed who tried to forecast where this trend will reach in any field.

Generation of power at frequencies higher than 60 c/s and voltages above 440 may be found necessary. Production of power by such methods as the fuel cell and thermo-electric generators may become practicable in the not too distant future, though much development work still remains to be done.

Nuclear propulsion of submarines has already been proved; but if it can be justified for surface warships it will present problems which have still to be tackled. The electrical engineer will be involved in the development of faultless and foolproof control systems, and in novel designs of auxiliary electrical machinery, capable of working continuously under very adverse conditions without risk of breakdown.

These and similar problems which are associated with missile installations will keep my successors in the naval electrical engineering field fully occupied, and will exercise not only their ingenuity, but also that of their colleagues in industry and elsewhere, whose enthusiastic co-operation will be just as essential in the future as it has been fruitful in the past.

In conclusion I wish to place on record my appreciation to the Board of Admiralty for permission to write this Address. I also wish to thank those authorities who so kindly provided illustrations and models, and in particular the Maritime Museum at Greenwich.

The Institution of Electrical Engineers
Abstract No. 3488
Jan. 1961

ELECTRONICS AND COMMUNICATIONS SECTION: CHAIRMAN'S ADDRESS

By T. B. D. TERRONI, B.Sc., Member.

'CHANNELLING—A SKETCH'

(ABSTRACT *of Address delivered 26th October*, 1960.)

In attempting to give some idea of the work carried out in industry by a telecommunication engineer concerned with carrier transmission, I propose to take a representative cross-section of his work and carry it through its stages of development to ultimate manufacture. The field is a large one, but there is one piece of equipment which can be said to be common to most eventual uses to which telecommunication will be put and this is the channelling, which is required for a multitude of purposes, whether these be allocated to speech, music or high-speed signalling. Channelling is intended to convey the translation of a speech band to its appropriate position in the frequency spectrum, so that an assembly of such channels can be handled as a unit block for grouping into larger blocks for subsequent transmission over the long-distance medium.

The manufacturer may have to supply communication equipment for use on open-wire lines, on cables or via radio links, and whichever of these bearer circuits is employed, the channelling equipment is common and the quality of transmission must conform to standards agreed by the C.C.I.T.T., whose main concern has been to ensure good quality over circuits up to 2 500 km in length, which may include international links. The C.C.I.T.T. has established that speech circuits shall have a controlled response over the band 300–3 400 c/s, so that speech channels of 4 kc/s nominal bandwidth can be assembled in blocks of 12 in the frequency range 60–108 kc/s. From Fig. 1 it will be seen that these basic groups are built into supergroups of 60 channels in the range 312–552 kc/s, which are next assembled in blocks of 300 channels to form a mastergroup in the range 812–2 044 kc/s, and these can form part of a supermaster group of 900 channels in the range 8 516–12 388 kc/s.

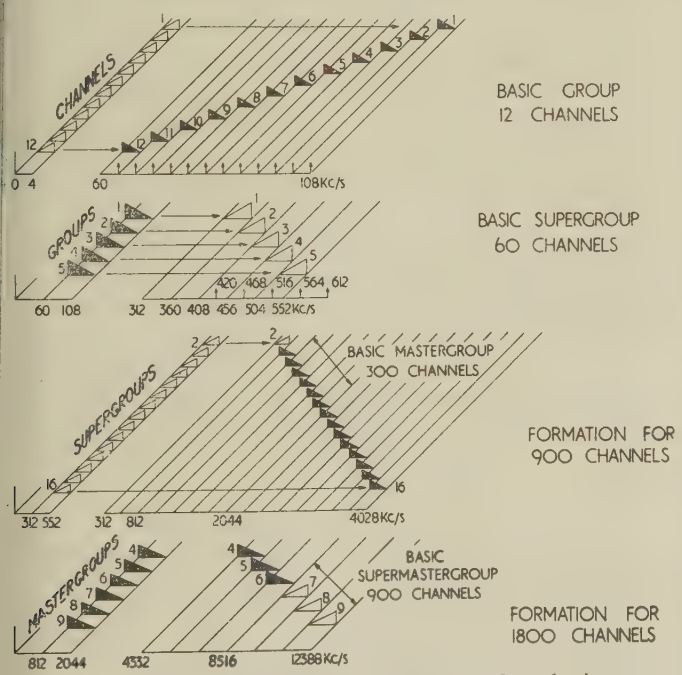


Fig. 1.—Build-up of channels into mastergroups from basic groups.

Since the formation of channels into basic groups entails the largest proportion of the equipment in a multi-channel carrier terminal, it is not unreasonable to give prominence to channelling and to consider initially some of the problems met with in building up the basic group of 12 channels in the band 60–108 kc/s.

The first method which naturally comes to mind is the direct one, which consists in taking the individual speech channels up to the respective bands 60–64 kc/s, 64–68 kc/s, and so on (with channel 1 occupying the position 104–108 kc/s), by modulation of channel 12 with a carrier frequency of 64 kc/s, channel 11 with 68 kc/s, etc. This modulation stage is followed by a highly selective filter which, in the case of channel 12, has to pass the wanted lower sideband, 60–64 kc/s, and reject the upper sideband, 64–68 kc/s.

This method was adopted in America and later in this country for use over the national telephone networks. At the time when it was introduced in America, some 25 years ago, magnetic materials of high quality were not available and a considerable effort was put into the design of resonators employing quartz crystals for building lattice-type filter networks which would conform to the stringent requirements for channel filters. To keep the crystal size small, for reasons of economy, it was necessary to make the translation frequency as high as practicable; on the other hand a high working frequency would mean high cable loss. With the basic block of 12 channels in the range 60–108 kc/s—which is conveniently inside the octave—the method lends itself by group modulation to provide a lower block of twelve channels in the range 12–60 kc/s, thus giving a convenient and economic band for transmission over cable without wasting too much of the low-frequency unused band, 0–12 kc/s—a considerable achievement in retrospect, which has stood the test of time. It has the merit of directness and simplicity in concept. Its success is dependent upon the design of filters of very high selectivity to eliminate the unwanted products of modulation to the desired extent—a characteristic calling for precision manufacture of quartz-bar resonators.

The alternative approach to the solution of this problem of frequency translation is to resort to double modulation, a technique first introduced in Germany some 20 years ago, no doubt partly as a result of the limited availability of quartz. This may be carried out in a number of ways, and a few of the more common methods employed throughout the world will now be considered. The problem is one of designing the important channel filter so that it operates at low frequencies for which good-quality components are comparatively easy to get. The lowest convenient carrier frequency for which design and manufacture of selective filters can be carried out is 8 kc/s, which will produce the two sidebands 4–8 kc/s and 8–12 kc/s. The upper band is eliminated by the filter and the 4–8 kc/s band is transmitted as a clean band which can now be modulated with 56 kc/s to give the required band, 60–64 kc/s, and an unwanted band, 48–52 kc/s—but now it will be noticed that the two sidebands

Mr. Terroni is with the Automatic Telephone and Electric Co., Ltd.

which have to be separated at 56 kc/s are well spaced so that a simple filter is all that is required to isolate the wanted band, 60–64 kc/s. This system was adopted by the French P.T.T. and has the merit that the first and difficult filter is the same for all channels. To the manufacturer this means a repetition job at the easiest part of the frequency range. This advantage, however, has been gained at the cost of an additional modulator per channel, with its associated filter, and one additional amplifier per group of 12 channels. Furthermore, 13 carrier frequencies are now required for modulation purposes instead of 12.

The double-modulation system adopted in Germany is based on a build-up from groups of 3 channels in the range 12–24 kc/s. This system requires only 4 subgroup modulators and it has the further advantage of requiring only 7 carrier frequencies. An alternative system standardized for use by the Services in this country and in America is based on subgroups of 4 channels in the range 4–20 kc/s.

The variants in double-modulation methods described above have been introduced for cheapness in manufacture which results largely from the best use being made of capacitors and adjustable ferrite-core inductors, and for flexibility in filter design to cater for the requirements laid down by different administrations in respect of overall performance under severe environmental conditions. This is particularly important, since most countries aim to use identical techniques for military and civil equipments.

The manufacturer, actively engaged in competing in the export market, is directly concerned with these different design methods, and he must therefore assess the relative merits in terms of ease of manufacture and grade of labour required, bearing in mind that as a result of contractual obligations he may be called upon to establish manufacture in a foreign country.

The differences reside essentially in the filters. In one case these are made up of specially ground quartz crystals which are assembled in lattice formation and require component matching and balancing. In the other case, coil-capacitor filters designed by the insertion-loss method give the required overall attenuation with an optimum number of elements.

Equipment Practice

Having described some of the electrical design features of telecommunication equipment, I will next consider the method to be employed in mounting the apparatus to give the desired qualities of low manufacturing cost and small volume without sacrifice of accessibility for maintenance. The practice is controlled to some extent by the need for compatibility with equipment already installed by different administrations, since this affects station layout, maintenance routines and power supplies.

If we examine the mechanical design features of communication equipment we find that in the past apparatus was mounted in a can to form a compact unit which could be conveniently tested. The can was fixed to a heavy steel panel, where the units were wired up to form a circuit entity such as an amplifier. The panel was fitted to a heavy angle-iron rack, where all panels were interconnected through rack cabling to provide a carrier repeater or a rack of audio repeaters. Improvements in design eventually led to the equipment known as Type 51 which was adopted by the Post Office a few years ago and is perhaps the neatest of design developments to date. It consists of an assembly of units in cans which can be individually clamped in any position between the rails of a framework by means of a fixing clip. The panel thus formed can now be pushed into the appropriate rack position, where connections with other panels can be picked up by plugs which bridge the sockets provided at the ends of the panel with the sockets fitted to the rack, now made of thin sheet steel. This design marked a considerable advance in equipment practice and it had an air of finality. To some extent this was conditioned by the need for valves and their location to provide adequate cooling and ease of replacement. However, with the introduction of transistors and parallel developments in the components field—more particularly ferrite materials which gave the prospect of much improved filter performance for smaller bulk and weight—the restrictions of the past could no longer be accepted. The solution arrived at can be approached in a simple and logical way by considering a box unit where the need for interleaving components imposed by space considerations has led to inaccessibility which can only be overcome by employing hinged box sides for mounting.

In the opened position the equivalent of mounting in one plane is apparant, and this is the ideal arrangement for assembly in manufacture and for subsequent maintenance. The next step, therefore, from this to card-mounted equipment with wiring on the underside, is an obvious one—so much so that almost identical and independent designs have appeared at about the same time in this country, in America and in Germany. With this arrangement, a multitude of mounting brackets requiring plating and finishing have been dispensed with, and components have become, to a large extent, self-supporting.

Methods of Dealing with Components

Coming now to the method of dealing with components, this would appear to be a matter of philosophy as much as of manufacturing adaptability. We might, for instance, decide to mount the equipment in hermetically sealed cans, thereby eliminating the effects of humidity on unprotected components. This makes for cheapness of component manufacture, but results in inaccessibility for inspection and maintenance; furthermore, sealing is an expensive process if it is to be effective. The alternative is to protect individual components by impregnation or sealing compounds. But here high temperature must be avoided in sealing, since this causes physical distortion or magnetic core ageing, resulting in instability which is particularly serious at carrier frequencies. A further difficulty is encountered in preventing the sealing compound gaining access to inductor windings, causing increased self-capacitance and instability under varying temperature conditions. A compromise in design is therefore desirable. One which is found convenient and economical in practice is to segregate critical high-frequency circuits such as filter networks from the rest of the components and to mount the filters in sealed cans. Whilst this entails the cost of drying prior to filter adjustment, and again drying prior to sealing, it does obviate the need for coil and capacitor processing. This brings in its train a considerable reduction in testing time in manufacture, since it will be appreciated that any protective coating applied to components such as coils and capacitors will result in a shift in the component value through alteration either of the physical constants or of the dielectric constants and this shift must be measured and allowed for in manufacture.

Environmental Testing

It is perhaps well to point out that the designer's work has not been completed when he has solved his problem of multiplexing. When the prototype design has been made up in its final form, it must be submitted as a system to a series of searching tests under conditions simulating extremes of climate and vibration, as well as low pressures which obtain when the equipment is flown to different parts of the world.

Thus the extremes of climate which have to be considered are those obtaining in arctic regions with temperatures down to -40° C, and in the tropics, where temperatures of $+50^\circ$ C may be reached with high humidity. Then, having found that the operation of the equipment has remained unimpaired by such

remes, and that the metal parts show no signs of rust, the nufacturer may have to submit the equipment to vibration by unting it in an army truck and running this at speed over a cially prepared boulder track.

Taking Stock

Having considered channelling of the present day in general ms, it may now be interesting if we take stock and re-examine position to see if we can improve our designs in the light of vances which have taken place in the components field. The sent approach to multiplexing is one which has been governed evolution based on a foundation established some 40 years . This foundation is, in fact, the derivation of multi-channel rking on a single-sideband basis by frequency division—since ailable wire circuits have a limited economic bandwidth. The jor problems in the design and manufacture of frequency vision multiplex (f.d.m.) equipment are those of distortion in plifiers, which has to be very low indeed to prevent inter-dulation effects generating noise in channels, and the attenua-n/frequency characteristics of filter networks, which have to designed with good impedance characteristics, low pass-range s and low distortion as well as high stop-band loss. These tworks are, in the main, critical in design as they have to ieve the stability requirements with time and over a wide nperature range and must be economical in components. The ssibility of eliminating these two requirements in the design of annelling equipment is well worthy of study, as it would nplify manufacture and testing and thereby reduce cost.

Time Sharing

Time sharing instead of wavelength sharing has been con-nplated for many years, but the interest has been limited in e past by the disability arising from the large number of rmionic valves required and by the need for a wide baseband handle the pulses. The modern transistor now, however, ings pulse modulation into prominence as a feasible proposi-n because of its small size, small power consumption and uivalence to valves for many applications.

Much has been written on pulse techniques, and I would erefore confine myself to pulse code modulation (p.c.m.), as has now been generally accepted as the most promising system transmission available to us. The aim of this new technique to send high-speed telegraph pulses, which represent speech, sampling the speech amplitudes at a frequency equal to at ast twice the highest speech frequency transmitted, i.e. 00 c/s. This has been established as being quite adequate for ithful reproduction. The instantaneous amplitudes thus deter-ined are transmitted in code as digits of a binary system. In actice, speech level variations of 60 dB, covering soft and loud lkers and other causes of level difference, are adequately catered r by a 7-digit system if these levels are distributed in a negative ponential manner by non-uniform quantization, that is to say larger number of steps are allocated to low speech levels and a aller number of steps cover the high levels. The coded square ulses having been produced (Fig. 2), the corners are rounded to duce the bandwidth by transmitting to line via a filter having a se approximation to a Gaussian loss characteristic with linear ase response. On the line the normal repeater of the frequency vision system is replaced by a regenerator which merely decides ether a pulse is incoming or not; and so long as the peak nplitude of the noise and crosstalk interference is not greater an half the amplitude of the pulse, a new undistorted pulse ill be transmitted to line—a scheme which is robust in concep-n and positive in action, with a high margin against false entification. This system will provide a signal very much more mmune to interference than that obtainable with orthodox f.d.m.

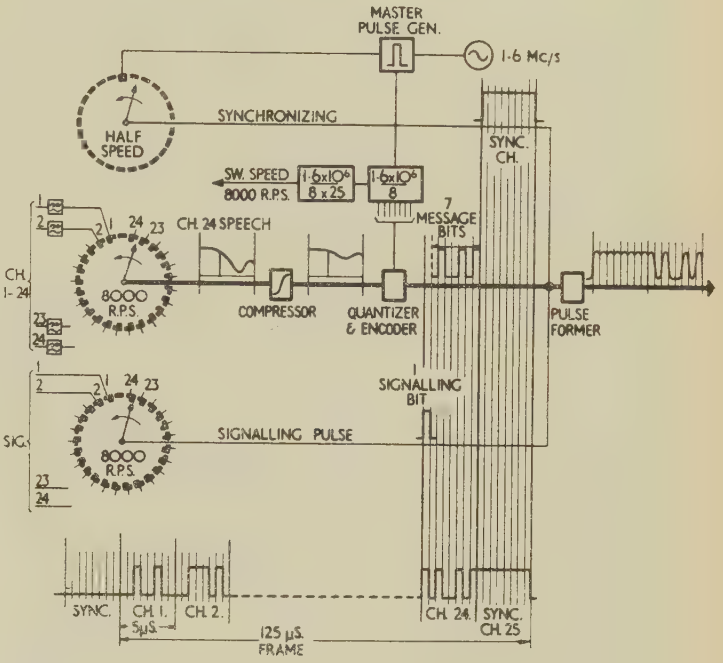


Fig. 2.—24-channel p.c.m. transmitting terminal. Showing the formation of signal impulses transmitted to line.

systems, and as a result, cable performance requirements become very much less exacting. The terminal equipment now takes the form of printed wiring boards fitted with small and inert components such as transistors, capacitors and resistors, and the few filters required become very simple indeed.

P.C.M. Application

Let us now examine the application of p.c.m. to the problem which is going to be the most pressing in the near future. The greatest proportion of circuit-miles in the United Kingdom and elsewhere is made up of multi-pair cables or junction cables up to 50 miles in length. The maximum congestion occurs in cables shorter than 20 miles in length, where the duct capacity in large cities like London and New York is rapidly being used up, and the need for some means of providing additional circuits economically over existing cables is calling for urgent considera-tion. If we consider a normal junction cable, which may have 20 lb/mile conductors laid up with paper insulation, we find that the crosstalk increases at 6 dB per octave, whilst the attenuation increases as the square root of the frequency. Furthermore, attempts to increase the channels in an audio cable by applying f.d.m. systems to the individual pairs is limited by the need to meet a crosstalk figure of 60 dB. With this in mind a low-cost short-haul carrier system was designed by the Post Office employ-ing frequencies of 24–108 kc/s in one direction and 158–222 kc/s in the other, to provide an additional 12 channels on 90% of the cable pairs. The far-end crosstalk for such a cable at 220 kc/s can be expected to be just better than 40 dB, but by employing compandors, the equivalent of a 60 dB crosstalk figure can be achieved.

With a p.c.m. system, however, although the bandwidth required is greater, considerable advantage is to be derived from the logarithmic exchange of bandwidth for signal/noise ratio in accordance with the Shannon law, with the result that much lower crosstalk margins can be worked to, provided the repeater spacing is correctly chosen.

With a number of parallel p.c.m. channels, the predominant

source of continuous noise will be due to near-end crosstalk because of the level difference obtaining between the 'go' and 'return' circuits; and with randomly distributed pulses over the parallel pairs, this noise might be expected to approximate to a Guassian distribution. In this case it can be shown that a peak-signal/r.m.s.-noise ratio of 21 dB will give rise to one error every half hour. This error probability of 1 in 10^8 is much more than necessary, and in practice the signal/noise ratio for p.c.m. does not need to be more than about 20 dB as against a 60 dB crosstalk requirement on normal carrier working.

A study of near-end crosstalk in audio cables shows that a 20 dB crosstalk margin can be exceeded up to a working frequency of 750 kc/s when the regenerative repeaters are spaced at 2 000 yd intervals. This frequency band will conveniently accommodate a 12-channel p.c.m. system on at least 90% of the cable pairs, employing a 7-digit binary code. In practice, network planning considerations may well decide on a 24-channel 1·6 Mc/s system employing 7 digits per sample plus one digit for signalling and a 25th channel for synchronizing. It should prove possible to apply this to 20% of the cable pairs with regenerators at 2 000 yd. Although the repeaters may appear to be very close together, it is to be remembered that these are of a simple transistorized type and quite small; they can be assembled in buried pots requiring very little power fed over the cable; and, of course, any synchronizing means employed can be common to all regenerative repeaters at the repeating point. This is to be compared with brick-built huts housing rack-mounted valve-type repeaters with frequency 'frogging' and heavy power requirements.

Cost comparisons lead one to believe that, whereas short-haul systems might be expected to improve in design to the point of showing economy over lengths of audio cable down to 24 miles, this distance could come down to 9–10 miles with p.c.m.

Economy with p.c.m. results from the large amount of common design in the equipment used in the terminals, with consequent repetition in manufacture. Complicated filters of different designs are replaced by diodes, resistors and capacitors, identical for the different channels and having wide tolerances.

When we come to apply p.c.m. to an existing network, it must be borne in mind that flexibility points will be required for the interconnection of carrier systems (f.d.m.) and p.c.m. or time-division systems. A possible basis would be to consider the 24-channel block as the new basic unit which requires transmission at 1·6 megabits/sec, or megabauds. This block can be transmitted over audio cables from minor exchanges to group centres, and this again can be extended over paired cables between group centres and zone centres by employing regenerators at 2 000 yd intervals. One might visualize a number of subscribers' television receivers operating over a paired cable at 1·6 megabauds. On the other hand, a 5 Mc/s television channel sampled at, say, 12 Mc/s would require 6 pairs of a cable (one pair for each of the digits), each carrying signalling at 12 megabauds and requiring regenerative repeaters to be spaced at 500–1 000 yd. This 'superbit group' of 12 megabauds woul also handle a 5 Mc/s multi-channel f.d.m. baseband, where th 8 digits required per sample could be transmitted over separa high-speed pairs making use of common timing equipment regenerative repeaters.

Thus a waveguide route might well carry a number of radi frequency carriers each conveying pulse trains up to near 200 megabauds assembled from 'superbit groups' of 12 meg bauds generated for a 5 Mc/s television channel, a 5 Mc/s f.d.m baseband and 7 interleaved 24-channel p.c.m. blocks.

It is interesting to note that the coaxial-cable system sho diminishing crosstalk with increasing frequency but increasi difficulties of regulation and equalization, so that p.c.m. mig be used to provide a more rugged and economic system usi about 160 megabauds and repeater spacings of 1 to 2 mil The radio link, on the other hand, shows little advantage throu using p.c.m. because of the unstable condition of the microwa propagation paths caused by fading. It may, however, all radio channels to be worked at closer spacing. This does n bring us to the end of our possibilities; in fact it might be sa that it is the least part of the application of pulse techniques practice, since these have already made considerable strides the switching field, where routing in an exchange, which normally carried out by electro-mechanical switches and re contacts, can now be visualized as the field proper to tim division multiplex techniques.

The conversion of electrical voice signals into time-divisi form is a costly process, but once the speech information h been produced in this form, it becomes practically and ec nomically desirable to derive the advantages of both transmissi and switching in one integrated arrangement which may w include wire pairs, coaxial cable, microwave link and circul waveguide. Electronic systems will provide faster service a better quality through the association of amplifiers with improved telephone instrument to take up losses and to oper a voice-frequency ringing system with key set instead of t present dial.

All these are interesting developments for the future, but th are still many problems to be solved before we can hope attain a fully digital world-wide communication network wh will provide uniformity in multiplexing, switching, and tra mission of high-speed digital information for the control industry, airports, and computer offices for commerce.

For many years the introduction of p.c.m. will be depend upon direct competition step by step with established syste which can claim the advantage of many years of developm to a high state of perfection and economy.

In the meantime, it is likely that p.c.m. will find early appli tion to increasing the capacity of junction cables; and this w make available, at the terminals, equipment which will be pres into use to extend time sharing to longer-distance circu thereby taking advantage of the benefits to be derived fr tandem operation of regenerative repeaters.

CENTRE, SUB-CENTRE AND GROUP CHAIRMEN'S ADDRESSES

The Institution of Electrical Engineers
Abstract No. 3492
Jan. 1961

MERSEY AND NORTH WALES CENTRE: CHAIRMAN'S ADDRESS

By DONALD A. PICKEN, Member.

'THE EFFECTS OF ELECTRICITY ON HUMAN BEINGS'

(ABSTRACT *of Address delivered at* LIVERPOOL *3rd October*, 1960.)

To understand the effects of electricity on human beings it is sirable to know something about the parts of the body which ly be affected.

The skeleton consists of a hinged structure controlled by a iscular system which operates entirely by contraction whether expanding the chest or gripping a tool. The muscles themves are, for our purpose, of two types: the striated muscle, uich is controlled by the central nervous system; and the rtially striated muscle, which has inbuilt nervous features and n therefore operate in isolation from the central nervous stem. The muscular system contracts on the application of lses of current at a frequency of about 100 per second. The gher the peak value of the pulse the more powerful the ntraction.

The nervous system consists of a network of what literally are nductors and serves to carry information and instructions from nsors to the brain and thence to the operating muscles or in e case of the partly striated muscle directly.

The brain consists of discreet centres all operating on an ctrochemical basis. Memory is stored probably by an nborate system of voltaic cells which operate like the electronic ain, except that in addition to operating in two states, charged uncharged, they can also operate by degrees of charge. emory also serves as the origin of the nervous pulses which ntrol most parts of the body.

The lungs consist essentially of air sacs in which air is brought to very close proximity to blood, with only a thin permeable embrane between and through which oxygen can pass from e air to the blood and carbon dioxide can be eliminated from e blood. These air sacs are contained within the thorax, which enclosed by the thoracic wall under the ribs and by the diaragm. Contraction of these muscles increases the volume of e thorax and this induces air into the lungs; relaxation of the uscles causes air to be expelled.

The blood circulatory system is operated by the heart, of which e right-hand system pumps blood through the lungs, where e products of oxidation of body tissues, principally carbon oxide, are eliminated and oxygen is absorbed into the blood, iich is then circulated by the left ventricle via the arteries and e capillaries of the muscles, skin, etc., where oxygen is supplied enable energy to be developed and the products of the reaction minated. The blood then returns via the venous system to e right auricle, and thence from the right ventricle to the lungs, nence it returns via the left auricle to the left ventricle. The art can be controlled by the central nervous system—the creased rate of beating caused by fear is an example—but a art separated from the body will continue to beat so long as urces of energy are provided and products of reaction removed.

Normal 50c/s alternating current can be thought of as 100 pulses per second, as the nervous system operates on a pulse irrespective of its polarity. The application of 50c/s alternating current to muscles will therefore affect them just as do the hundred pulses per second of the nervous system.

If electrodes are applied to the hand and forearm so that current passes through the wrist, effects can be studied without risk to life; a current through the wrist of 0·5–2mA will cause a slight sensation and a current of 3–6mA will affect the nervous system in such a way that pain in the region of the wrist, such as that caused by pulling out a hair or prodding with a needle, will not be felt.

If the current is increased greater difficulty will be experienced in operating the muscles, and at 8–20mA the muscles will become immobile by complete contraction and it will not be possible to release the grip of the hand. The sensory nerves will still be insensitive, and pain apart from heat due to the resistance at the points of entry and exit will be slight. If a current of 20mA or more were suddenly applied the contraction could be such as to crush bones or tear muscles.

If, instead of through the wrist, current is passed, say, from hand to foot at about 20–50mA, the muscles of the thorax will be forced into contraction, the lungs will be abnormally filled with air and will remain so as long as current flows, and as no exhalation can take place, waste products will build up in the blood; in a few minutes irreversible changes will take place in the brain which will prove fatal. On cessation of the current the patient may not recommence breathing, owing to muscular fatigue, non-elimination of combustion products or effects of current on nerve ends. These effects are normally reversible, and artificial respiration will be effective if promptly applied and maintained.

If the hand-to-foot current is of the order of 50–500mA the current density through the heart will be of the same order as that through the wrist which affects the nervous system of the wrist, and instead of the heart beating in a co-ordinated manner the individual fibrils will contact haphazardly; the resulting phenomenon, known as 'fibrillation', will almost entirely stop blood circulation. Even short-time passages of current (0·1sec) can have this effect if they occur during the sensitive phase of the heat beat, and all periods of current flow exceeding one heart cycle (about 1sec) must include such a phase and will produce fibrillation; for when the heart fibrillates it does not normally resume the sensitive phase so that co-ordinated beating is not resumed even when current ceases. Unless effective steps are promptly taken the brain will suffer such changes that death is inevitable.

To restore heart action, either a severe shock must be given to the heart to produce tetany of the muscles or chemical methods of stopping the fibrillation such as injection of acetylcholin must be adopted. There is doubt as to whether fibrillation can be eliminated spontaneously, but this in any case is only of academic

Mr. Picken is H.M. Electrical Inspector of Factories.

interest unless oxygenated blood circulation to the brain can be restored within 3–5 min.

If a current of the order of amperes flows, there will be the contraction of the thorax and the current through the heart will be such as to cause the heart muscles to contract, but when the current ceases, provided it has not been too prolonged, the heart is likely to resume beating either spontaneously or as a result of movement such as results from a blow on the chest or the action of normal artificial respiratory techniques. Thus higher currents are less likely to produce immediate death than currents under one ampere, and indeed currents of the order of amperes may cause fibrillation to cease, and were it not for other effects would be a useful method for lay treatment of fibrillation. Such currents, however, as well as giving serious risk of burning may also have effects on the nervous system which render it incapable of controlling the muscular system. The risks of these effects is such that lay application of these countershock treatments for fibrillation is dubious, particularly as the diagnosis of fibrillation is difficult for a doctor without electrocardio apparatus.

Wherever such currents flow there will be internal heating effects which may be very serious, and whilst unlikely to cause instant death, may well do so after a few days unless steps are taken to eliminate the myoglobin which may result from the injuries. The drinking of alkaline liquids such as one dessert-spoonful of sodium bicarbonate in a tumbler of water is very beneficial for such a purpose.

The resistance of the body has two components, both variable. One of these is skin resistance, which may vary from 10^4 ohms/cm^2 to 10 ohms/cm^2, depending chiefly on the humidity of the skin. This factor is influenced by perspiration, which can result from the passage of a small current through the skin, so that, whatever the initial dryness may be, with a 240-volt supply the current flow will be such that, unless the contact can be broken immediately, perspiration will ensue and the current will increase and set up an effect which will rapidly reduce the skin resistance to a negligible value.

The second effect is the internal resistance of the body, which is of the order of 200–800 ohms and varies inversely with the applied voltage.

The overall resistance varies inversely as the 1·5–1·9 power of the voltage. Thus reducing the voltage from 240 to 110 will increase the effective resistance by about 4 times and decrease the current by about 8 times.

This is borne out by a statistical survey carried out by W. H. O. de Bats, Senior Electrical Inspector for Holland, who found that after eliminating factors not common to both systems the 220-volt system was 6·5 times as likely to produce electric-shock accidents as an 119-volt system.

Voltages of the order of 20–60 volts a.c. are likely, given effective contacts, to result in inability to let go if one of the contacts is gripped as an electric tool would be.

Similarly a voltage of 40–100 volts across a path which traverses the chest will stop the breathing, and 80–600 volts is likely to result in fibrillation. Over this range of voltage, nerve blocks and burns are the most probable effects, but if the contacts are poor, for example through somewhat insulatory floors or footwear, the voltage across the body may be 80–600 volts wit much higher system voltages.

Capacitor and lightning discharges normally are too short t affect the nervous system, but serious burns can result, thoug they are often merely superficial.

In a recent case the discharge of a 114 μF capacitor, charge at 4 kV, across a man's chest had no effect apart from seve burns and nervous shock, and the injured man was able to wa to the ambulance room for treatment.

Frequency has little effect at any practicable supply valu though increasing the frequency by 10 times may increase th safe values by 20%, i.e. a typical let-go current would be 15 m At 10 kc/s the let-go current is about 80 mA, and above abo 15 kc/s there is little effect on the muscular system, thoug heating effects are unchanged.

Direct current only affects the muscular system with the on and cessation of the flow; these can cause a violent jerk, and th sensation resulting when contact is broken no doubt accou for the impression that 'd.c. throws you off'. Small direct curren up to 0·1 amp are negligible unless protracted, but higher curre and small currents for a longer period, in addition to causin serious burns, can give rise to electrolytic effects which ma seriously affect the nervous system.

The effect of ambient temperature as a factor in electric sho is shown by the following table, which gives a twenty-ye (1928–48) month-by-month list of accidents which have occur in Switzerland. It was prepared by the Power System Inspecti Service of the Association Suisse des Electriciens:

January	137
February	131
March	182
April	138
May	218
June	305
July	301
August	262
September	219
October	212
November	163
December	141

It will be noted that the incidence rate increases broadly as t temperature increases and subsequently decreases again. Bea ing in mind that the hottest months are also the holiday mont when the incidence is likely to fall, these figures are quite strikin Less elaborately based figures for ten years in a section England show a similar tendency:

January	48
February	35
March	55
April	60
May	62
June	60
July	56
August	67
September	63
October	60
November	66
December	45

The Institution of Electrical Engineers
Abstract No. 3459
Jan. 1961

SOUTHERN CENTRE: CHAIRMAN'S ADDRESS

By R. GOFORD, Member.

'THE TELECOMMUNICATION ENGINEER AND THE BACKGROUND IN WHICH HE WORKS'

(ABSTRACT *of Address delivered at* PORTSMOUTH *5th October*, 1960.)

The Civil Service

The present Civil Service structure has existed since the plementation of the Northcote–Trevelyan Report in 1870. It ludes about 1·1 million officers, amongst whom Post Office nnicians and officers up to the rank of assistant engineer are arded as departmental classes, and scientists, engineers and er professional grades are included in the Treasury classes. ecutive engineer and higher engineering ranks are departntal variants of the works group of professional and technical sses.

The Post Office

A Committee under Lord Bridgeman considered the organizan of the Post Office in 1932 and determined that executive ntrol should be decentralized from Headquarters to a regional anization with the widest possible devolved powers; secondly t there should be close local co-ordination of all aspects of telephone service; and further that Headquarters were to nsider and determine policy, and regional organizations to be ncerned with execution of policy with as little hindrance as ssible. Sir Thomas Gardner was given the task of implenting the report and proposed the creation of eight provincial ions, each containing a number of telephone areas with phone managers at their head in control of engineering, ffic, sales and accounting functions. They were to be given nsiderable devolved powers. The regional system was in gh working order by 1940 and contributed materially to the ort the Post Office was called upon to make during the war.

The Present Organization

In 1950 a committee under Mr. D. O. Lumley reported on present system of regionalization and made recommendans the following year.

The Postmaster General with the Assistant Postmaster General eaks for the Department in Parliament, is Chairman of the st Office Board and presides at panels of the Post Office visory Council. The Director General is permanent head, Vice-Chairman of the Board and is assisted by three Deputy rectors General and the Engineer-in-Chief. He is the account- officer for the Post Office vote. At the next level are directors ponsible for posts, telecommunications, establishments and anization, personnel, mechanization and buildings, radio vices, finance and accounts, savings, research and two Deputy gineers-in-Chief. There are other officers controlling ancillary partments. The ten regions have each a director, one or two puties and five controllers. Under their control are 67 teleone managers, of whom nine are in London, and 470 head stmasters. The telephone managers control 6313 telephone changes, of which 5392 are automatic and 921 manual. Of a ff of 338000, 0·05% are at Headquarters, 10% in London partments and 90% approximately in regions. There are o Whitley Councils, Engineering and Non-Engineering, with mbers nominated by the Department and the trade unions. addition, there is a network of joint production committees ding up to the Joint Production Council.

The chief regional engineer is the principal technical adviser to the regional director in all engineering matters. The area engineer is responsible for supervising engineering works under the technical control of the chief regional engineer and also for exercising the main engineering powers devolved on the area under the telephone manager.

Devolution is practised and encouraged to the greatest practicable extent, but the parliamentary system requires that the Postmaster General shall be accountable to Parliament for the manner in which he spends funds allotted to him. The Accountant General's Department collates estimates built up from basic information furnished by telephone managers and head postmasters via the regional directors, and these are examined by a series of standing committees of which the Comptroller and Accountant General is a member. In addition, audits of telephone managers' and head postmasters' accounts are carried out yearly by officers of the Comptroller and Accountant General or the regional director.

The Status of the Post Office

A White Paper, CMD.989, was presented by the Postmaster General in March, 1960. This announced that the Post Office would be encouraged to conduct its business as a commercial enterprise and to balance income and expenditure. Commercial accounts were to govern financial policy, and the Post Office would contribute a basic sum of £5 million per annum to the Exchequer in lieu of income tax. The Post Office would be free from the provision in the budget of self-balancing revenue and expenditure. All Post Office receipts would be paid into a trading fund from which all Post Office payments would be made. Parliamentary control would be maintained by authorization of borrowing of new capital, and amongst other measures by the examination of commercial accounts by the Comptroller and Accountant General and the Public Accounts Committee. Treasury control would be applied to investment as for the public sector generally.

The Engineering Department

The Engineer-in-Chief is responsible to the Director General for all Post Office engineering work, and this supervision is partly specialist and partly technical. It covers research, development, design, maintenance, construction, control of performance, training, submarine cables and ships, provision of radio stations, testing and inspection of materials and equipment, and promotion of senior engineering grades.

The Regional Organization and Engineering Branch

The Engineering Branch in collaboration with the Telecommunications Branch is responsible for the engineering development and maintenance of telecommunication services, and, in conjunction with the postal branch, for development and maintenance of engineering equipment used in the postal services.

Telephone Area Organization

Telephone areas differ widely in size, density, nature of territory and extent of plant. There may be from one to four area engineers, and in the densest areas a deputy telephone

Mr. Goford is with the General Post Office.

manager. London has eight areas co-ordinated from the London Telecommunications Regional Office. The telephone manager is responsible for maintaining satisfactory service and co-ordinating engineering, sales, traffic, accounting, commercial and personnel aspects of the work.

The Engineering Division

Areas vary from 8 to 14 000 sq miles in size and from a staff of 500 to more than 3 000. Functional working is favoured at the level at which engineering responsibility splits: at area engineer level where the area contains more than one; at the next lower grade where there is only one area engineer. The control system is applied as far as possible in installation of telephones and minor works, major works, and maintenance.

The works control operates the works programme, feeds stores, tools and transport to the jobs, deploys working parties to the best advantage and progresses works against the programme. The maintenance control receives fault reports and issues them to maintenance staff for clearance. It keeps maintenance records and statistical data.

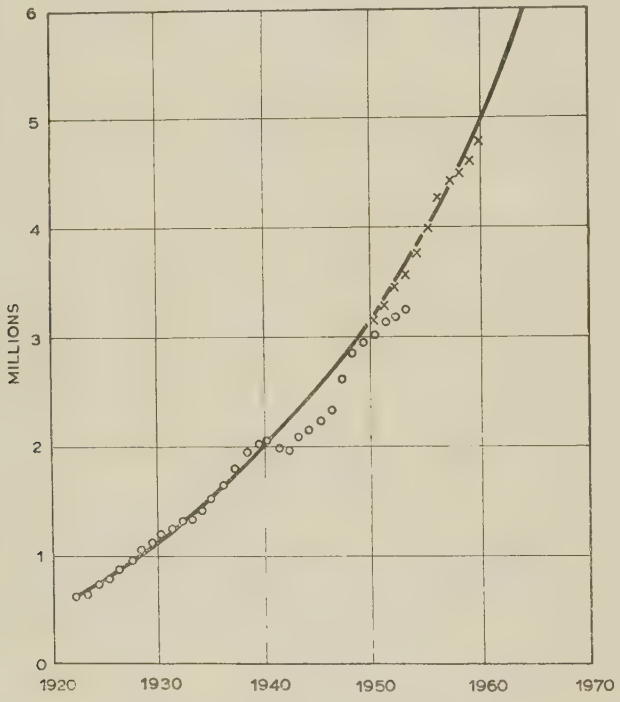


Fig. 1.—Telephone growth.
○ Number of exchange lines up to 1950.
× Number of exchange connections from 1950.

Growth of the Post Office

The Post Office is an immense trading organization. The public spend over £400 million per year on its services. The amount of money handled by the Post Office per annum exceeds £5 000 million. Fig. 1 shows the growth of the system from the early days of the telephone to date. A normal growth curve appears to be followed, and the demand appears to be still on the lower part of such a curve; i.e. it can be expected to steepen for some time before the growth commences to fall away again. The effect of the last war is apparent, and so also is that of the fairly recent tariff changes. Demand is rising again, and it seems possible that normal growth will again be achieved although the evidence available at present is not clear enough for one to assert that this is so.

Fig. 2 shows the variation of gross demand for telephones in a typical area and indicates quite clearly how would-be subscribers respond to all domestic, political and economic events that affect the public generally. This makes estimation of demand difficult.

Telephone Area Functions

The principal preoccupation of the Engineering Division is the provision of adequate plant to maintain a good service to the subscriber. The Engineering and Traffic Divisions determine traffic loading on routes between exchanges, and into and out of the area for periods of one, three and five years ahead. The Sales Division estimates the number of subscribers who will require service in the next five, ten, fifteen and twenty years. The Engineering Division translate these figures into equipment and external cable quantities and plan and execute the work of installing them, the larger works being undertaken by contract and controlled from regional headquarters with liaison and work supervision by the telephone manager's staff.

A new exchange site is chosen by the External Planning Group so that external plant costs may be a minimum. The Ministry of Works buys the site; the Post Office determines the shape and size of building to accommodate the equipment required. The Ministry of Works designs and controls the erection of the building. The equipment of smaller exchanges and minor extensions are carried out by the telephone manager's staff. Traffic flow through exchanges is studied, and the interconnection of equipment is flexible to enable adjustments to be made to smooth out the incidence of traffic at various switching stages.

Other functions of the Post Office include the tracing

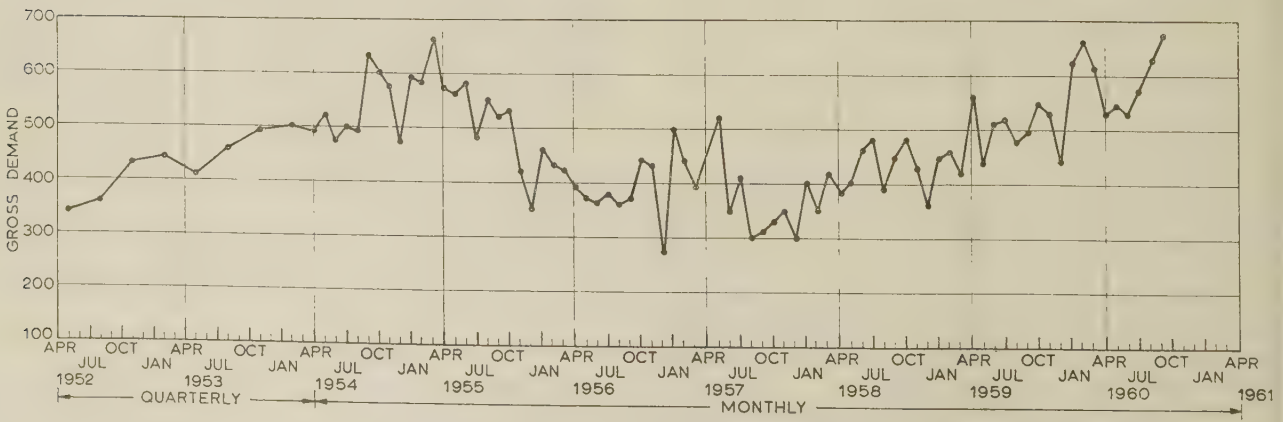


Fig. 2.—Gross demand for telephones in Portsmouth area.

dio interference, inspection of amateur radio equipment, stallation and maintenance of Post Office services such as ctric light and power, heating, ventilation, postal aids, etc., intenance of plant, renewals to counter plant deterioration, stallation of telephones and private branch exchanges, etc. ny demand may be made upon it at short notice, and emergencies may face local officers in isolated circumstances. This situation was often experienced by the staff during the war, and most Post Office employees feel some satisfaction that most other services and undertakings depend upon them and are never disappointed at the Post Office's reaction to the most exacting demands made upon it.

The Institution of Electrical Engineers
Abstract No. 3476
Jan. 1961

RUGBY SUB-CENTRE: CHAIRMAN'S ADDRESS

By E. S. HALL, Member.

'THE NEW LOOK IN INDUSTRIAL ELECTRONICS'

(ABSTRACT *of Address delivered at* RUGBY 12*th October*, 1960.)

The Electronics Industry

Originally the term 'electronics' was almost self-explanatory; ctronic apparatus always contained devices known as elec- nic valves, depending for their operation on the movement electrons in a vacuum or of electrons and positive ions in a s. With the advent of semiconductors and magnetic ampli- rs, however, electronics must now be regarded as the con- uing development of the science and techniques originally ablished around the use of the electronic valve. The term dustrial' covers the commercial application of electronics, cluding domestic radio and television and its associated trans- ssion equipment, telephones, electronic instruments, and all litary or defence projects.

During the last six years the output of the electronics industry this country doubled, whereas the output for the electrical lustry as a whole increased by only 1·6 times. Relatively, the nount of industrial electronics activity is small; the industry s been dependent upon Government expenditure on defence ojects.

Industrially the situation on development manpower is equally sturbing. The White Paper, 'Scientific and Engineering Man- wer in Great Britain, 1959' shows that between 40% and 50% the qualified electronics engineers in this country are on fence projects. Much greater effort is required on industrial ork.

The New Look in Old Applications

Some of the earliest applications of electronics in industry, ch as photo-electric relay devices, welding controls, motor eed controls and generator voltage regulators, are still in mand to-day. Considerable developments have taken place, ising largely from the availability of new components, new lves and photocells and the development of the magnetic nplifier.

The photo-electric relay was introduced to industry over years ago; it used the vacuum caesium photo-emissive type cell. Since the war gas-filled and photo-multiplier types and miconductor light-sensitive devices have been used; recently e cadmium-sulphide photo-conductive cell has been introduced. fra-red photo-sensitive devices and photo-voltaic cells con- rting solar energy into electrical energy will no doubt in the ture find their use in industry, but their cost at the present ne is high.

Welding controls for resistance welding were one of the earliest plications of electronics to industry. The electronic control nsists basically of a timer controlling a magnetic or electronic vitch. The development of the ignitron with the mercury-pool cathode extended the switch rating, and digital counting valves have superseded the original resistance-capacitance timing.

Electronic speed control of d.c. motors has found extended usage since the war. *IR* compensation can readily be applied so that a flat speed/torque response can be obtained, and feed- back is used for stability. For precision drives or applications requiring frequent reversal of the motor rotation, electronic control is used in conjunction with a Ward Leonard combina- tion of machines—an arrangement particularly suitable for multiple drives where a number of shafts have to be accurately controlled relative to each other.

Electronic voltage regulators were originally developed 25 or 30 years ago; their general acceptance in industry has occurred during the last 15 or 20 years. High accuracy of regulation, quick response and high stability can be obtained with absence of moving parts. The regulator consists of three essential circuit-elements, namely the voltage-sensitive circuit, an error- signal amplifier and a power rectifier supplying the field of the exciter.

Magnetic amplifiers have been introduced into regulators within the last ten years, electronic valves being replaced by non-linear magnetic-core devices. For the larger ratings a sub- exciter is used owing to the limited power which can conveniently be obtained from the magnetic amplifier.

Magnetic amplifiers have many other uses. Their main advantages are:

No moving parts.
No starting delay.
Input electrically isolated from output.
Control circuits electrically isolated from one another so that several inputs can be used.
Control circuits of low impedance.
Robustness and ability to withstand vibration.
Increased reliability.

The dynamic amplification (power gain for one-second time- constant) of a self-saturating amplifier on a 50 c/s supply can be as great as 100 000. To reduce size, supply frequencies of 400 c/s have often been used and 3 000 c/s is not unknown.

The future of the magnetic amplifier will undoubtedly be affected by the advent of transistors and the semiconductor rectifier.

New Developments in Industrial Electronics

Radar was one of the outstanding electronic inventions of the Second World War. On the conclusion of hostilities, attention was given to the commercial and industrial applications of this equipment; in consequence it has since been applied to ground control for airports, commercial air navigation and cloud detection.

Navigational radar has become virtually standard equipment

Mr. Hall is with Associated Electrical Industries Ltd. (Electronic Apparatus vision).

on merchant ships, the United Kingdom being the world's principal supplier to shipping lines of all nationalities. The latest form of marine navigation radar includes means of presenting a 'ground-stabilized' p.p.i. display, so that land masses, buoys and other fixed objects appear as stationary echoes on the screen.

Industrial television has now become generally accepted. For example, the use of television in a slabbing mill has enabled the transfer of ingots from the shuttle car to the ingot buggy to be speeded up, and the difficult operation of feeding the mill with hot steel quickly enough to keep it working to capacity has been considerably assisted.

The generally accepted standard used in this country is a horizontal resolution of 600–650 lines, with interlacing. The weight of the camera unit has been reduced to 7–10 lb, and attachments such as remote focus, remote lens change and similar facilities are available. Colour television has also been used in some applications, but for general industrial use the extra complication and cost is not usually justified.

Electronic computers, both digital and analogue, are finding increasing application in industry. Analogue computers are widely used in the aircraft and the chemical industries for the study of complex processes; also by electrical engineers to study stability problems on servo mechanisms, and the behaviour of complex transmission and distribution systems. Digital computers for data processing, stores control, accounting, banking, wages and other clerical processes are becoming generally accepted in industry.

Special-purpose digital computers are also used for the numerical control of machine tools. The original information from the drawing is scheduled and fed into the main computer in the form of paper tape; the computer calculates the correct path to be followed by the cutter of the machine to produce the required finished dimensions. This information is in the form of a magnetic tape which is fed into the shop-floor controller, which then controls the operation of the machine.

Static switching units to replace relays and contactors are under active development; the advent of the transistor has made this possible. Typical logic functions of these units are known as 'and', 'or', 'memory' and 'nor' units. The combination of these logic units to carry out a complex switching operation requires a knowledge of Boolean algebra, and the resulting equipment may be more costly; nevertheless, it is anticipated that the improved reliability will frequently justify the expense.

Machine-tool controls include speed-control systems, automatic co-ordinate setting equipment, steering control and tracer control.

The use of electronic speed control on a machine tool frequently simplifies the drive, sometimes avoiding the necessity of a change-speed gear-box.

Electronic automatic co-ordinate setting equipment has been applied to horizontal and vertical boring machines and j borers. In most equipment the co-ordinates can be set manually or automatically from punched card or tape. C small quantities its use avoids the manufacture of elaborate jig

Tracer controls have been developed of the single-dimension two-dimensional and three-dimensional types for use wi copying lathes, milling and boring machines. The work-pie is reproduced from a model made in wood or other eas worked material.

Steering controls are designed to allow the machine operat to steer the cutter over the face of the work piece quickly a accurately from a hand-wheel located at a convenient positic

Improvements in Manufacturing Techniques and Product Appearance

Associated with the continuing technical advance in electroni there has been considerable effort directed towards improving t manufacturing methods and the appearance of the products.

Perhaps the most revolutionary change is the introduction the printed-circuit technique, which, however, is not yet fu proven for industrial use. In this system the conventio insulated wiring is replaced by copper conductors which a printed on an insulation board to which they are attached, a the components are soldered direct to the conductors. Ot improved manufacturing processes include wrapped connecti to avoid the use of solder, and potting techniques where co plete assemblies are encapsulated, thereby preventing contami tion by the atmosphere.

It must be agreed that technical performance is of grea importance than appearance for most of the industrial appli tions of electronics. Nevertheless, an attractive looking prod not only has sales appeal but encourages accuracy in the use the equipment and leads to greater overall efficiency.

The Future of Industrial Electronics

What of the future? New developments are proceed apace. The transistor and the controlled semiconductor rec fier may eventually make the electronic valve obsolete. Co ponents are becoming smaller and more reliable. Developme in micro-miniaturization will undoubtedly result in saving weight and space, should further increase the reliability of equipment in service, and ultimately result in a saving in cost

Automatic control of processes using on-line computers undoubtedly become more widespread. Electronic control machine tools and other production machinery will continue grow, and its use will undoubtedly be extended to ancill processes such as inspection and handling.

Computers and other electronic equipment are likely revolutionize present systems of banking and accounting, a more and more clerical processes will become automatic.

The Institution of Electrical Engin
Abstract No. 3
Jan. 1

NORTH-WESTERN UTILIZATION GROUP: CHAIRMAN'S ADDRESS

By C. AYERS, B.Sc.(Eng.), A.M.I.C.E., A.M.I.Mech.E., Associate Member.

'SOME EFFECTS OF AUTOMATION'

(ABSTRACT *of Address delivered at* MANCHESTER 11*th October*, 1960.)

It has been apparent for some time that the word 'automation' has become of great importance in engineering circles, and it is evident that its use requires some qualification. In many men's minds the term is synonymous with mechanization, this is not entirely true.

The process of mechanization follows a basic pattern in p

Mr. Ayers is with the English Electric Co., Ltd.

lly every application. Initially there exist several basic and crete operations. These are joined by suitable links to form ntinuous chain of events or a complete process. At appro- ate points in the chain an appreciation of the performance is ained to allow human intervention as necessary for correc- n, and at this stage the process of mechanization ends. uld the appreciation of the process be made to control the cess an automatic or automated plant results.

Thus progress in automation does not involve any basically / concepts but is more accurately defined as a new attitude mind. Taking this broad view a little farther it may be said t automation is the best integrated use of resources, both men materials, to produce a desired result; but this appears to be basic definition of 'engineering'. One may say, therefore, t no definition of 'automation' is possible, but that only gress towards that end can be established.

Apart from the technical problems attendant upon the applica- n of this thought to any process, several other problems have e solved.

The question of economics is always complex and is often ly balanced, being dependent upon such factors as the cost the ultimate end-product, the return on capital investment, and the attitude of mind adopted by the user. However, it appears to be generally the case that physical effort can be replaced by an automatic process, but that the replacement of mental effort presents a more difficult problem particularly if concerned with decision making.

The integration of man himself into the automation picture demands skill and patience when dealing with such facets as retraining, possible redundancy, and payment by results which in some cases is the only system with which an operator is familiar. A change of attitude towards work on the part of the operator is thus necessary.

The expanding boundaries of technology have presented educational problems in that specialization is now necessary at an early stage in training. While this is necessary in order that technology shall not stand still, it does not appear to be conducive to the development of engineers in the broad sense, who will be required to conceive and carry to fruition schemes of automation which by their very nature demand the close integration of several skills.

While what has been said applies to the industrial environment, it is equally applicable to the office, where there is much that could benefit from an application of the principles of automation.

A NOTE ON THE FORMULA FOR LOSS IN A FERRITE

By L. Lewin, Associate Member.

(*Communication received 3rd September*, 1960.)

The properties of a ferrite are usually expressed in terms of the meability tensor

$$\begin{bmatrix} \mu & -j\kappa & 0 \\ j\kappa & \mu & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

ere μ and κ may be considered either as empirically determined ameters or as quantities to be calculated from material perties. In particular, the combinations $\mu \pm \kappa$, which refer the propagation of circular polarization, can be calcu- ed, in the loss-free case, in terms of the gyromagnetic ratio $= -2{\cdot}8$ Mc/s per gauss), the saturation magnetization M_s d the applied magnetostatic field H. The calculations* rt from the equation $dM/dt = \gamma M \times H$, and lead to

$$\mu \pm \kappa = 1 + \frac{\gamma M_s}{\gamma H \mp \omega} \quad . \quad . \quad . \quad . \quad (1)$$

order to introduce the effects of loss, an empirical damping m is added to the equation of motion. Different forms have en suggested, but they all lead to the same structure of final mula. Hogan uses

$$dM/dt = \gamma M \times H - \gamma\alpha[M \times (M \times H)]/|M|,$$

which leads, after some manipulation, to

$$\mu \pm \kappa = 1 + \frac{\gamma M_s(1 \pm j\alpha)}{\gamma H(1 \pm j\alpha) \mp \omega} \quad . \quad . \quad . \quad (2)$$

The original expressions are appreciably more complicated than this, and only reduce to the relatively simple form (2) when the combination $\mu \pm \kappa$ is taken.

The purpose of this note is to show an extremely simple relation between (2) and (1), whereby it is possible to go from the lossless form to the lossy form by a simple change of variable.

If both numerator and denominator in (2) are multiplied by $1 \mp j\alpha$ it takes the form

$$1 + \frac{\gamma' M_s}{\gamma'(H + j\beta) \mp \omega} \quad . \quad . \quad . \quad . \quad (3)$$

with $$\gamma' = \gamma(1 + \alpha^2)$$

and $$\beta = \omega\alpha/\gamma' \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Hence the lossy case is obtained from (1) by increasing the gyromagnetic ratio by the factor $(1 + \alpha^2)$ and adding an imaginary term to H to represent the damping. This transformation is far from obvious from the original form of the equations. One simple consequence might be noted. Although the factor $1 + \alpha^2$ is not much different from unity for most ferrites, especially those with a narrow line width, direct deductions from resonance experiments determine γ' rather than γ, and the difference may be significant if accurate g-values, for example, are being sought.

* E.g. see Hogan, C. L.; 'The Microwave Gyrator', *Bell System Technical Journal*, 2, 31, p. 1.

Mr. Lewin is with Standard Telecommunication Laboratories, Ltd.

621.313.334 : 539.12

The Institution of Electrical Engineers
Paper No. 3336 I
Jan. 196

THE PHASE CONTROL OF ROTORS AT HIGH SPINNING SPEEDS

By P. A. EGELSTAFF, B.Sc., Ph.D., H. J. HAY, M.Sc., Ph.D., G. HOLT, J. F. RAFFLE, M.Sc., Ph.D., and J. R. PICKLES, A.M.I.Mech.E.

(*The paper was first received* 17*th March, and in revised form* 15*th July*, 1960.)

SUMMARY

A complex motor (called the spinning head) is described which can drive 20 lb rotors at speeds up to 60 000 r.p.m. and at the same time allow the phase relative to another unit or a standard to be controlled. The system has a natural phase stability of 0·5°, and with a simple servo system, phase control to 0·1° is possible.

The unit was developed for neutron spectrometer work and its value in this field is discussed. A number of spinning heads are in use in several neutron spectrometers and have been shown to be very reliable in operation.

(1) INTRODUCTION

(1.1) The Purpose of the Spinning Head

The instrument to be described was developed to drive neutron 'choppers' for time-of-flight spectrometers. Essentially, the technique used with these spectrometers is to modulate a continuous beam of neutrons from a reactor by means of a high-speed rotating shutter (called a neutron chopper) and subsequently to time the neutrons over a given flight path. The distribution in arrival times at a detector at the end of the flight path gives the velocity spectrum of the neutrons in the beam. Spectrometers of this type have a wide use in neutron physics and reactor physics experiments.[1]

The design of the chopper to produce the initial short burst of neutrons can take many forms depending on the particular range of neutron energies of interest and on the mechanical properties of the materials to be used. In the paper the discussion is restricted to the type of chopper consisting of a rotor, made from material of high neutron-stopping power, through which a set of diametric slots have been cut so that a neutron burst is produced each time a slot is in line with the neutron beam. This is shown diagrammatically in Fig. 1(*a*). The neutron path referred to axes fixed in the rotor is an Archimedean spiral and the velocity bandwidth and mean velocity transmitted are dependent on the size of the entrance and exit apertures, the shape of the slot walls, the diameter and the rotational speed. Present chopper designs are restricted by the fact that these factors are not independently variable.

The diameter of the chopper is determined by the condition that the neutron burst must be sharply defined and the intensity in the 'off' position must be low. For some experiments this means a diameter of 20–30 in. The aperture of the slit must be large in order to achieve a high neutron intensity in the pulse and thus a high peripheral velocity is required to keep the pulse short.

The development of the phase-controllable units to be described allows the conventional single chopper to be replaced by two or more choppers in series [Fig. 1(*b*)]. This phased system has several advantages which outweigh the corresponding increase in complexity over the single-chopper system.

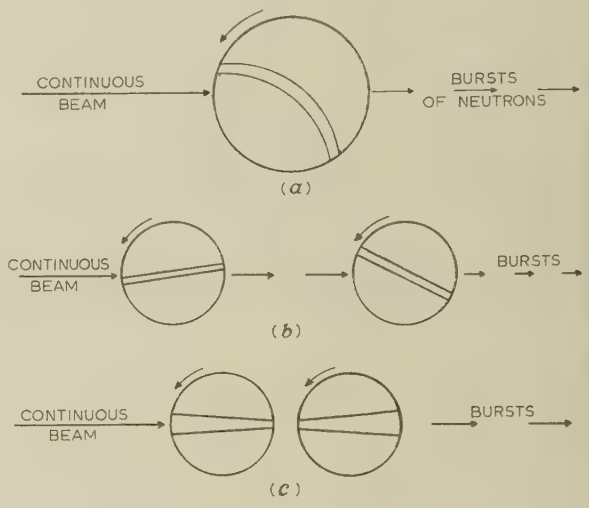


Fig. 1.—Single and phased chopper systems.

First, there is an advantage in greater safety.* If the sing chopper of Fig. 1(*a*) is replaced by two smaller ones, the store energy is greatly reduced; in fact the small chopper can, i general, be completely shielded against the release of the who of the stored energy whereas large single rotors cannot be s shielded. Secondly, there is an advantage of flexibility in desig Two small choppers arranged in series, with variable separatio and phase difference, are equivalent to one large rotor wit variable diameter and slot curvature. Used in this way th phased-chopper system acts as a mechanical monochromato for neutrons. In addition, the background in an experime can be determined simply by rephasing the choppers so tha the second one is closed before the first one opens and the pulse neutron beam is then cut off. The addition of a third choppe to this system can allow the pulse repetition frequency to b fixed independently. Thirdly, there are special advantages fo the fast-neutron chopper.[2] This type uses a very diverge neutron beam and the slits have the geometry shown in Fig. 1(*c*) The use of two choppers here has the advantages, compared wit a single chopper, of halving the neutron pulse length (since th tips of the choppers are going in opposite directions) and allowing the effective slit width to be altered by variation of th relative phase.

The degree of phase control required depends upon the pa ticular experiment, but, in general, the upper limit is about and the lower limit about 0·05°.

Finally, the use of phased-chopper systems implies the use small-diameter choppers and consequently high rotational speed if the peripheral velocity is to be kept high. In addition, some experiments the property of high angular velocity valuable in itself. Consequently the spinning head was designe with both phase control and high rotational speeds in mind.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Messrs. Egelstaff, Hay and Holt are, and Mr. Raffle was formerly, at the U.K.A.E.A. Atomic Energy Research Establishment. Mr. Raffle is now with the Cementation Co., and Mr. Pickles is with the Rover Co. Ltd.

* A very high degree of safety is essential, since these instruments operate in are where many people may be working, and are placed in proximity to the core o reactor which is both an expensive and a potentially hazardous unit.

(1.2) Techniques for Phase Control at High Rotational Speed

reliminary work on the spinning head was concerned with assessment of the merits of three methods of phasing control. first of these methods was purely mechanical with the drive he various choppers connected by spring-loaded gear trains. s method is reliable since the phase relationships are positively ned for all speeds and conditions, but the complexity of the em becomes very great if 0·1° accuracy is required, together n flexibility in operation (i.e. variable path and phase between units). In addition, the power absorbed in the gear system essitates a primary drive consisting of either a mains-driven tor with high step-up gears or a high-speed high-power motor n special power supplies. For these reasons the gear-nected system appeared likely to be complicated and some-at inflexible in practice and so further work on it was ndoned in favour of electrical phase control.

here are two methods of electrical phase control, one olving the use of induction or d.c. motors and the other ng synchronous motors. The first of these methods achieves se control essentially through accurate speed control employ-an elaborate servo system, since, in normal circumstances, the ed and phase of these motors are indeterminate. The second thod takes advantage of the fact that several synchronous tors driven from a common supply run at the same average ed and, within limits, have fixed phase relationships. This thod therefore offers the possibility of phase control with (at st) a simple servo system, it being assumed that phase jitter these motors is due to fluctuations in the load which, with table care, could be made quite small. Thus, of the two trical phase-control methods, the synchronous-motor method ms to be the more attractive, but, compared with the availa-ty of d.c. and induction motors, very few suitable synchronous tors are manufactured. Consequently, two preliminary periments were carried out on the available synchronous tors.

n the first experiment two 50 c/s mains-driven motors running ball races were set up and their natural phase stability was died under no-load conditions. It was found that the relative ase of the two motors was stable to $\pm 0{\cdot}05°$ over short periods, ich supports the idea that phase jitter in synchronous motors due to fluctuations in load.

The second test was carried out on 3-phase hysteresis motors.[3] is type of synchronous motor appears to be the only one ich is easily started and is adapted to high-speed operation. nlike a conventional motor, the poles on the rotor of the steresis motor are printed on by the driving field after the tor has reached synchronous speed. The purpose of the test s to examine whether the printed poles were stable, and the ase difference between two such motors running in ball races 400 c/s was measured under various conditions. It was found at the printed poles were very stable. For example, if one otor was loaded so that a lag of up to 20° was obtained, the otor would return to its original phase relative to the other otor when the load was removed, even after running under ad for several days. The natural stability in these tests was und to be poor because of the lower power (about 1/10th at of the 50 c/s motors employed in the first test) and the low rtia of the motors.

As a result of these tests it was decided to base the spinning ad on the 3-phase hysteresis motor and to construct it in such way that the energy losses due to bearing loads, etc., would kept to the order of 10 watts, and the fluctuations in the losses ould be small. With choppers of 5 to 20 lb weight this would ep the lag angle nearly constant, and hence the natural phase stability should be in the range 0·1–1° and, with simple servo control, stability of 0·1° should be achieved.

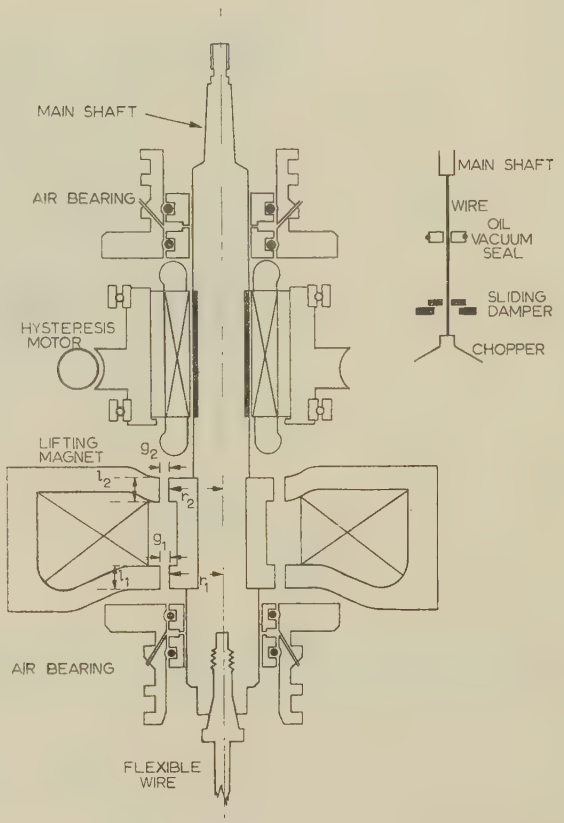


Fig. 2.—Cross-sectional diagram of the spinning head.

(1.3) Description and Operation of the Spinning Head

A cross-sectional diagram of the spinning head is shown in Fig. 2. The inset indicates the system of connecting to the choppers, which are arranged to spin in a vacuum chamber to eliminate windage losses and are attached to the main driving shaft in the spinning head by a $\frac{3}{32}$ in diameter steel wire. This wire passes through an oil-lubricated journal bearing acting as a vacuum seal, so that the spinning head itself can be at atmospheric pressure. The chopper is decoupled from the driving shaft by the wire so that any unbalance, due to inaccuracies in manufacture or uneven expansion of the rotor at speed, does not throw large loads on to the bearings. The main driving shaft rotates in journal air bearings and the weight of the whole system is taken on a magnetic thrust bearing. The use of gas bearings and the magnetic bearing ensures that the power losses are quite small and, in addition, the rotating parts are completely isolated so that the unit operates without significant noise or vibration. The components of the head are shown in Fig. 3.

The driving motor is a small 60-watt hysteresis motor which can produce an accelerating torque of 0·02 lb-ft and a synchronous pull-out torque of half this value. It is driven by an audio-frequency power amplifier fed from a crystal-controlled variable-frequency oscillator (see Section 2.3), and the synchronous speed is held to better than 1 part in 10^5. The acceleration time of the chopper is long (60–90 min to 60 000 r.p.m.), but this is not serious for a machine which is to operate continuously for many hundreds of hours at constant speed. The low input power is a disadvantage in that the chopper takes a long time to accelerate through shaft whirls or regions of unstable motion of the system. A

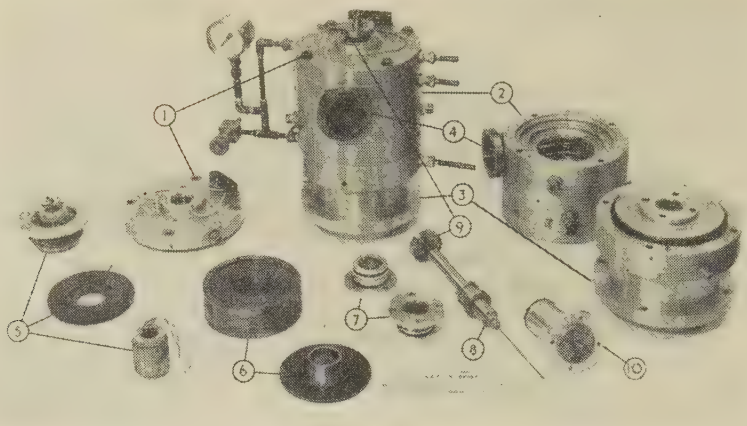


Fig. 3.—An assembled spinning head and its components.

1. Top plate with air bearing.
2. Housing for hysteresis motor and lifting magnet.
3. Housing for lower air bearing, oil seal, and damper.
4. Motor mounting for turning the stator coil.
5. Rotatable stator.
6. Lifting magnet.
7. Air bearings.
8. Shaft assembly, consisting of magnetic drum, upper air bearing surface, hysteresis rotor, lifting-magnet plunger, lower air bearing surface and flexible steel wire.
9. Magnetic drum.
10. Oil seal and damper assembly.

sliding-disc damper is incorporated to remove energy from any motion which causes the wire to move more than about 0·001 in from its normal position. At high spinning speeds the rotor axis remains fixed relative to its surroundings to better than 0·000 5 in and the thin shaft flexes synchronously to take up the rotor unbalance. Various types of torsional oscillation occur, which can lead to errors in phase; fortunately these can be controlled quite easily, as explained in Section 3.

When the hysteresis motor drops into synchronism, the phase of slots in the chopper relative to the driving-flux vector is arbitrary and the correct phase position must be set up. The time at which the chopper slots are parallel to the neutron beam may be determined from the pulse formed in a neutron or γ-ray detector placed behind the chopper, or in a previously aligned photocell which receives a beam of light reflected by a mirror attached to the top of the shaft. The pulses from this source, together with those from another chopper or a standard reference, are displayed on a cathode-ray tube, and a coarse setting of phase is made by reducing the power to the motor until it just slips from synchronism and then allowing it to slip to the correct position. The fine setting of phase is accomplished by rotating the stator of the hysteresis motor, and for this purpose the stator is fixed to a worm gear which is operated by a remotely controlled motor.

Because of the low power losses the deceleration time is very long and special deceleration methods are employed. One method is to pass direct current through two of the three phases of the hysteresis motor, when deceleration torques of about 0·02 lb-ft are obtained. Alternatively, deceleration is achieved by reducing the driving frequency in steps to zero.

(2) DESIGN AND PERFORMANCE

(2.1) The Thrust Bearing

The heaviest rotor planned for the spectrometers was of 10 in diameter and weighed about 20 lb. The thrust bearing to carry this load is a magnet which takes the form of a double-gap axial plunger, as shown in the general arrangement of Fig. 2. With such a magnet the axial force for small displacements may be written, using the notation of Fig. 2, as

$$F_a = F_{a1} + F_{a2} = 1{\cdot}1 \times 10^{-7}\left(\frac{F}{S}\right)^2\left(\frac{l_1 r_1}{g_1^2}S_1^2 + \frac{l_2 r_2}{g_2^2}S_2^2\right)\text{lb} \quad . \quad (1)$$

in which S_1 and S_2 are the reluctances of the gaps and S is t total reluctance, F is the ampere-turns supplied by the coil, a the iron is assumed to be unsaturated. This formula has t correct theoretical form and the scaling constant has be determined experimentally to take into account the effec fringing fields. The value of F_a obtained from eqn. (1) cor sponds to about half of the maximum attainable lift with (measured) plunger depression, d, approximately equal to t gap width. These conditions have been found to give a sufficie operating margin.

Such a bearing is unstable, however, against sideways for arising because of eccentric gaps caused by machining erro by misalignment of the core in the yoke, or by freedom sideways movement in the bearings. The magnitude of th sideways forces is given by

$$F_{s1} + F_{s2} = \frac{2}{d}(e_1 F_{a1} + e_2 F_{a2}) \quad . \quad . \quad . \quad .$$

when the eccentricity e is small compared with the gap, a again the numerical constant allows for the experimental co ditions. Examination of eqns. (1) and (2) shows that the ra of sideways force to axial force is decreased by making g an as large as possible.

The distribution of the sideways forces along the shaft is su that the lower bearing must be made stronger than the upper o because the magnet is on the lower half of the shaft. Furth to keep the reaction on the lower bearing small, F_{s1} should kept less than F_{s2}.

This magnet requires 2 300 ampere-turns to lift 20 lb, and w a 2 400-turn coil of resistance 30 ohms the power dissipatio less than 40 watts. Thus the operating temperature of the c is not high. With a depression of 0·06 in the total sidew force per 0·001 in eccentricity would be about $\frac{3}{4}$ lb. In t system, 0·002 in is allowed for misalignment and machin errors in the magnet, and 0·001 in for misalignment in the bearing and mounting. This can lead to sideways forces 2 lb, which must be balanced by the journal air bearings stable operation. The magnet will start to saturate at the cen of the core as the flux density approaches 15 000 gauss, whi for this magnet, will be at 2 400 ampere-turns, about the va required to lift 20 lb. This is seen in the measured performa shown in Fig. 4. The inset demonstrates the increase in sidew

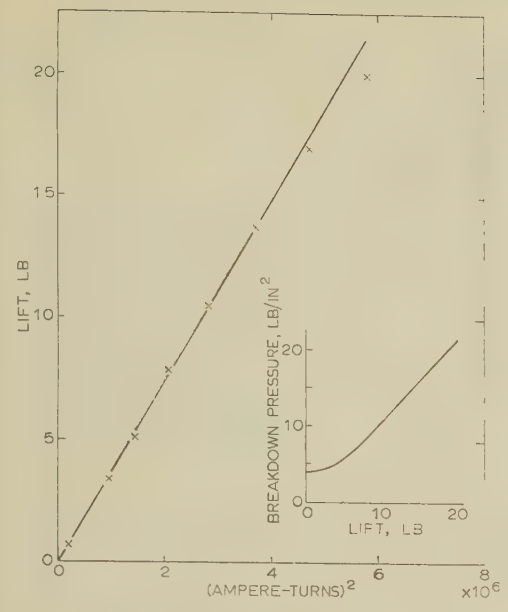


Fig. 4.—Lifting-magnet performance.

The numerical coefficient quoted in the text is the mean of the initial slopes for several magnets. The slope decreases as saturation is approached at high ampere-turns.

The inset shows the minimum air pressure required to prevent the bearings breaking down owing to the sideways forces associated with the axial load.

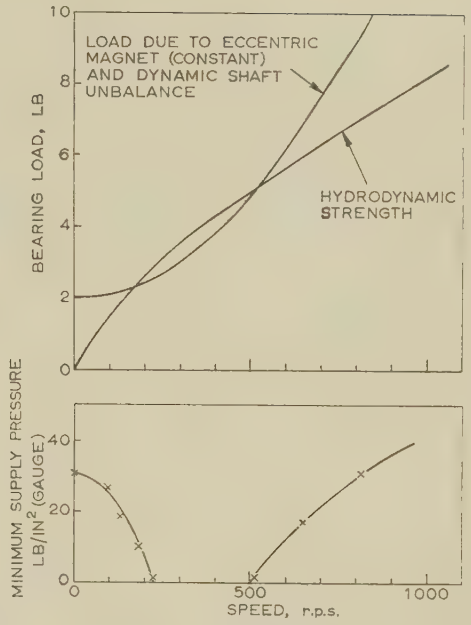


Fig. 5.—Calculated and observed loads on the gas bearings.

The minimum supply pressure must give a bearing strength equal to the difference between the total load and the dynamic strength. The dynamic strength was calculated from Reference 6 using $n = 0{\cdot}7$.

ce with lift according to eqn. (2), since the sideways force can estimated from the minimum air pressure required to prevent air bearings from breaking down.

A cylindrically symmetrical thrust bearing of this type will ve very low power loss when rotating. Eccentricities and omogeneities will lead to eddy-current losses dependent on square of the rotational speed, but these sources of loss quite small and it has not been possible to make an independent measurement of their magnitude.

When the current to the magnet coil is removed, the shaft ops to a ball race mounted just below the connection between $\frac{3}{32}$ in wire and the shaft.

(2.2) The Journal Air-Bearings

The loads which must be carried by the journal bearings arise m two causes, namely the sideways forces in the magnet and dynamic unbalance from the rotating shaft. The magnet n produce loads of about 2 lb, and although the rotating shaft balanced to better than 10^{-5} lb-ft, dynamic loads of up to o can be imposed on the bearing at rotational speeds of 00 r.p.s.

The bearing adopted is a pressure-fed air bearing with an nular gap of 0·001 in, and air can be supplied at pressures of up 80 lb/in^2 through six equally-spaced holes of 0·022 in diameter nd the centre of the bearing. Both bearings are 0·8 in long h the top bearing diameter 0·778 in and the bottom, 1·000 in. om the work of Shires[4] and Robinson[5] the strength of each aring at 30 lb/in^2 supply pressure should be 2 lb, which reases with supply pressure to 10 lb at 80 lb/in^2. The bearing hole diameter was made equal to that employed in various gas arings already in use. However, the work of the above hors show that 0·010 in diameter holes should increase the ength of these bearings for supply pressures up to 50 lb/in^2, d there is some experimental evidence in support of this.

The dimensions of the bearings are such as to give hydrodynamic properties,[6] i.e. when the bearings are rotating they l keep themselves supplied with air and can carry a load without the applied external pressure. The way in which the hydrodynamic strength increases with speed is shown for typical operating conditions in Fig. 5. Over part of the operating range the hydrodynamic strength exceeds the journal loads, and it would be expected that the external supply could be shut off over this range. That this is so can be seen in the lower part of the Figure, in which the measured variation of the breakdown pressure with speed is given; the breakdown pressure is measured by reducing the supply pressure until metallic contact is heard in the bearing. The uncertainties in the calculation are too great to predict the operation of the bearing precisely. Not all the units employed have shown the self-acting region, but all show a minimum in the breakdown-pressure/speed curve. Supply pressures of 40 lb/in^2 are usually more than sufficient to prevent breakdown over the whole range. At this pressure the air consumption is roughly 1 ft^3/min.

The bearings were mounted in rubber O-rings to assist in damping shaft vibrations and to make them self-aligning. The degree of clearance and the hardness of the rings were chosen carefully to damp adequately the shaft vibrations but not to allow too much free movement which would increase the sideways forces arising from the magnet.

The air supplies are filtered so that no particles greater than 5 microns in size can reach the bearings. In the event of air-supply failure, a large reservoir or standby compressor supplies air until the choppers have been decelerated. A low gas consumption is an advantage from this point of view.

(2.3) The Driving System

The driving motor is a 2-pole 60-watt 400 c/s hysteresis motor designed to give a synchronous torque at 400 c/s of 0·01 lb-ft. Measurements on the motor showed that it was possible to run at up to 1 000 c/s with little loss of torque by the use of higher voltages. Under these conditions the temperature rises, estimated from the resistance of the windings, were about 50–60° C.

The cobalt-steel rotor of the motor was mounted on the main driving shaft (Fig. 2) and ground to size. The stator was

mounted in ball races and by means of a worm gear can be rotated about the shaft, the current to it being taken in through slip-rings and carbon brushes. On starting up, the rotor comes into synchronism in an arbitrary phase position and the methods of setting the correct phase have been outlined in Section 1.3.

High-stability speed control is achieved by using a 100 kc/s crystal in the primary oscillator. Frequency-dividing circuits give outputs of periods 40, 200, 1 000 and 4 000 microsec, and by combining multiples of these a square wave is obtained with a period (variable in 40 microsec steps) between 240 and 20 000 microsec. This is filtered to give a sine-wave output at the desired operating frequency.

The 3-phase supply for the motor is developed from the single-phase sinusoidal output in the manner shown in Fig. 6. The

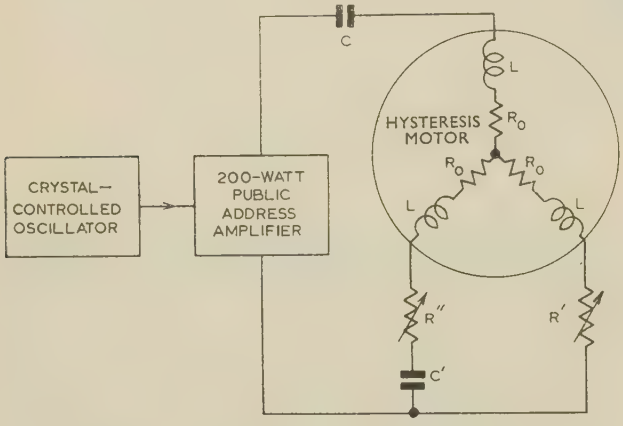


Fig. 6.—Driving circuit for hysteresis motors.

capacitor C′ is switched manually to follow the law $C' = 1/2\omega^2 L$, and the equal resistors R′ and R″ are adjusted with frequency so that arc tan $\omega L/(R_0 + R')$ is maintained at 60°. In this way the currents in each branch are kept of equal amplitude and 120° apart. In practice the phase separation need not be exactly 120° and R″ may be regarded as a trimmer to improve the balance. C' varies over the range 100–0·5 μF from 50 to 800 c/s, and ideally C should be maintained at twice the value of C' to keep the power factor near unity. However, satisfactory operation is obtained with C kept constant at 12 μF. During acceleration the frequency is increased in steps as the speed goes up and the current balance is corrected at the same time.

(2.4) Power Losses of the Whole System

The power losses in the rotating system are shown in Fig. 7, where the loss is plotted as a function of speed for various conditions of operation. The measurements were made by observing the free deceleration rate when the electrical connections to the stator were open-circuited.

Curve (*a*) shows the losses from the main driving shaft alone without the flexible shaft attached and represents the losses in the air bearings, windage losses on the shaft, and any loss in the magnetic bearing. The curve is significantly less steep than the square-law expected from simple theory of laminar flow, possibly owing to the effect of the rubber mountings. Curve (*b*) shows the power losses when the $\frac{3}{32}$in diameter flexible shaft has been added, together with the oil seal and the damper. After several hundred hours' running at high speed, curve (*b*) reduces to curve (*c*), the maximum losses being about 10 watts.

The addition of choppers to the end of the shaft gives a further increase in the power loss due to the flexing of the shaft doing

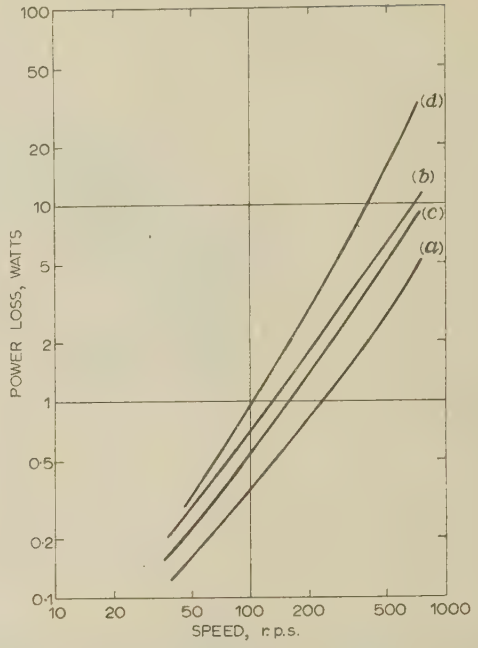


Fig. 7.—Observed power losses in the spinning head.

(*a*) Main shaft only.
(*b*) Main shaft plus flexible shaft, oil seal and damper.
(*c*) As (*b*), after running in.
(*d*) Typical losses when a chopper is added.

work in the oil seal and dampers; a typical case is shown curve (*d*). The total power loss is still low enough for o purposes, and the fluctuations in the losses are not significant.

If a rotor should go out of balance at speed owing to so uneven stretching or the breaking off of some part, the con quent flexing of the shaft would do considerable work in t damper. This could readily exceed the driving power availa and cause the rotor to slow down, so forming a type of built-i safety mechanism.

(3) TORSIONAL OSCILLATION AND PHASE CONTRO

The spinning head was designed to have very low and a rather uniform power losses, since this assists the phase cont by ensuring that the lag angle of the synchronous motor is m or less constant. Because of this it is expected that rand fluctuations in phase are rather small, and the main problem phase control lies in damping the natural torsional oscillatio of the system. The driving system should therefore be c which allows all spinning heads to use the same power supp and which, in addition, couples the heads together so that if external disturbance causes oscillations they are in phase in heads. It has been found that the commonest source of c turbance is fluctuations in the mains voltage.

Various methods of driving the hysteresis motor have b tried and are outlined in Table 1.

The substantial improvement in phase stability in going fr method (i) to method (iv) tends to confirm that the spinn head itself does not contribute much to fluctuations in ph and that these originate in the driving system. When us methods (iii) and (iv) very little long-term drift in phase encountered, and the errors in phase which are seen are due the excitation of the natural vibrations of the system. T nature of these vibrations is described in Section 3.1, followed the methods employed to damp their amplitude to l than $\pm 0{\cdot}1°$.

Table 1

PHASE STABILITY UNDER VARIOUS DRIVING CONDITIONS

Driving method	Natural phase stability of one head with respect to another	
	Over a few hours	Over days
) 3-phase oscillator to three amplifiers and then to motors	±1·5°	Bad
) Single-phase oscillator, one amplifier to each motor, and phase splitting as in Fig. 6	±1°	Several degrees
) Single-phase oscillator, one amplifier to a pair of motors, each motor with phase splitting as in Fig. 6	±0·5°	±0·5°
) Single-phase oscillator, one amplifier and phase-splitting circuit to a pair of motors whose windings are joined together	±0·3°	±0·3°

(3.1) Natural Torsional Frequencies

he torsional arrangement of the apparatus is shown in 8. A chopper (of $\frac{1}{4}$–1 lb-ft^2 moment of inertia) is mounted

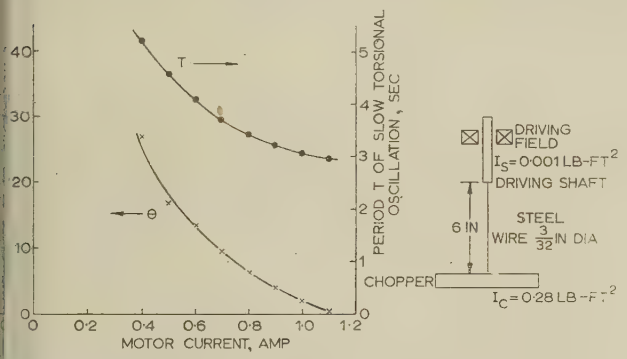


Fig. 8.—Synchronous characteristics of the spinning system.
The zero for the lag angle is arbitrary at 1·1 amp.

a 6 in long $\frac{3}{32}$ in diameter wire suspended from a driving ft (of inertia 1/1 000 lb-ft^2) which is weakly coupled to the ring-flux vector.

uch a system has four natural frequencies, as may be seen considering the mechanical analogue of a mass joined by a spring to another mass which is itself suspended on a spring. Torsional oscillations were excited in the rotating system by modulating the driving power with a varying frequency, and the values of this frequency for which a maximum amplitude was obtained were noted.

The most important torsional frequency is that of the shaft and chopper relative to the driving-flux vector. This is observed by measuring the oscillation of a light pulse reflected from the shaft relative to a signal from the oscillator, and is found to be $\frac{1}{3}$–$\frac{1}{10}$ c/s. Of secondary importance is the torsional oscillation of the shaft relative to the chopper, which is found to be 34 c/s and is observed by measuring the oscillation of a light pulse reflected from a mirror on the shaft relative to a light pulse reflected from a mirror on the chopper. The oscillations of the shaft relative to the driving flux and of the chopper relative to the shaft have been observed but are of little importance, the frequencies being about 5 and 2 c/s respectively. The measured torsional frequencies agree quite well with those calculated from the constants of the system.

The oscillation of the whole system relative to the driving flux is the most important because it is very easily excited and has a very low frequency. The actual frequency depends not only on the moment of inertia of the chopper but also on the strength of the coupling to the flux vector, which depends on the power supplied to the motor. Fig. 8 shows the typical variations in the period T of the torsional oscillation and the lag θ of the shaft relative to the driving flux as functions of the motor current. The latter relation provides one method of controlling the phase of the choppers.

(3.2) Torsional Damping

For many neutron experiments the natural phase stability of ±0·3° obtained with driving method (iv) of Table 1 is adequate. However, in some cases stability of ±0·1° is required and here it is necessary to provide damping for oscillations of frequencies $\frac{1}{3}$ and 34 c/s.

The 34 c/s frequency is not a source of trouble in practice since it is found to be well damped and the bass cut-off of the amplifier may be raised so that the motor does not get any power at this frequency. If the motor is disturbed, either by electrical or mechanical means, then the amplitude of the torsional oscillations at $\frac{1}{3}$ c/s is about 10 times that at 34 c/s. Consequently the desired level of phase stability may be achieved by providing damping for the $\frac{1}{3}$ c/s oscillations only.

A number of damping circuits have been employed with the various driving methods listed in Table 1. The common feature has been the generation of a signal, dependent on the lag angle, which is allowed to modulate the power to the motor at such a

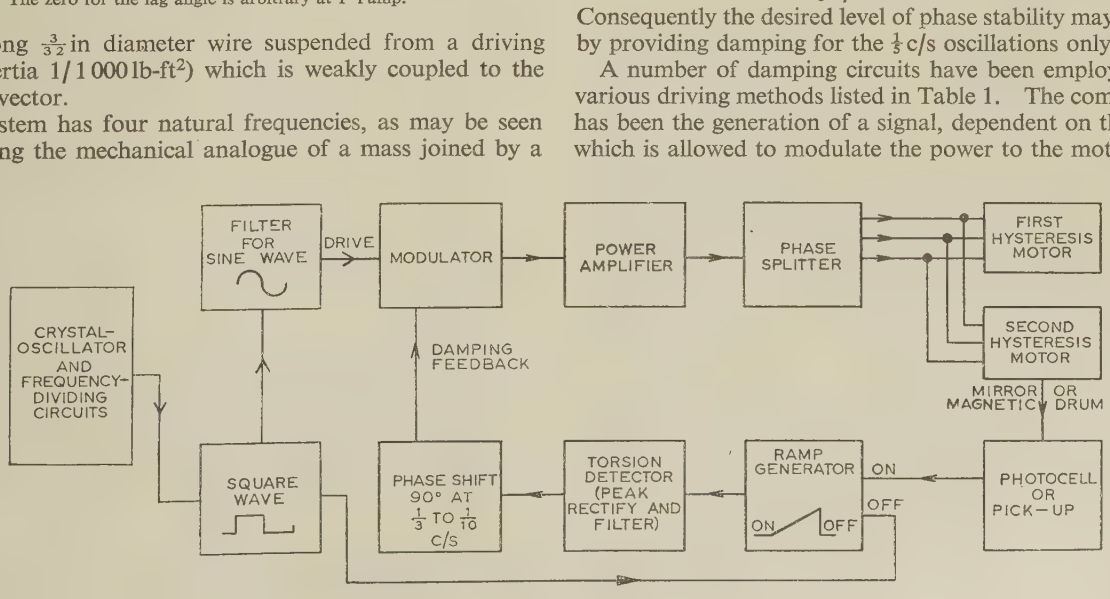


Fig. 9.—Block diagram for damping torsional swings of phase.

phase that the $\frac{1}{3}$ c/s oscillation is damped. Of the methods used, the most convenient is employed with driving method (iv) and is outlined in Fig. 9. A signal at the torsional frequency is generated from the time interval between a light signal from one chopper and a reference pulse from the driving oscillator. This is shifted in phase by about 90° so that it is proportional to the rate of change of lag angle, and is applied as negative feedback to the power amplifier. In this method, damping signals are received from one head only, the second head following the first through the coupling introduced by joining the motor windings together. Phase stability of the order of $\pm 0{\cdot}03°$ has been obtained by this method.

(3.3) Long-Term Stability

The natural phase stability over periods of several days has been given in Table 1. When torsional damping is employed, a stability of better than $\pm 0{\cdot}1°$ is obtained over several days. Consequently, long-term drift may be corrected manually, making adjustments of the stator position of one head every day or so. Long-term servo controls are therefore not needed, although two methods have been tried out. In one, the stator position is servo-controlled, so eliminating the need for manual adjustment, and in the other (only suitable if separate amplifiers are employed) the lag angle is servo-controlled by making automatic adjustments to the amplifier output.

(4) SAFETY MEASURES AND OPERATIONAL EXPERIENCE

The need for a high degree of safety has already been discussed, and, since the rotors themselves are run inside steel tanks whose walls are thick enough to withstand an explosion of the rotor, the ultimate safety of personnel and of the reactor is assured. However, it is desirable to prevent incidents if possible and consequently the spinning head has been fitted with devices which will disconnect the 3-phase supply and apply d.c. braking if anything unusual occurs. The aim of these devices is to sense any departure from steady-state conditions; because the spinning head is required to run for many days under fixed conditions, any departure from a steady state will indicate a fault of some kind.

The devices at present incorporated detect significant departures from the normal values of the speed of revolution, height of the shaft in the magnetic thrust bearing, vibration of the spinning head, temperature of the head, vacuum maintained in the tank and the performance of the vacuum pumps, air-supply pressure, and the oil supply for the vacuum seals. Failure of the a.c. mains or large variations in the lifting-magnet current will also act as a trip. The d.c. brake supplies are obtained from a trickle-charged battery to assure reliability, and the air supply to the gas bearings during deceleration is assured by building into the system either a large reservoir or a small reservoir and emergency compressor. It has been found in practice that the occurrence of a trip is a rare event, which indicates that the above quantities are well defined and that the spinning head operates reliably.

The recording of data from the neutron experiments must be temporarily stopped if the relative phase is disturbed by an abnormal event. The light signals can be used in a phase-sensitive detector which disconnects the neutron counters from the time sorter until conditions return to normal.

Machines using these spinning heads to obtain phase contro are now in operation at two reactors. At the N.R.U. reacto at Chalk River, Canada, an experiment to study the scatterin of thermal neutrons by moderating materials employs fou phased rotors, at rotational speeds up to 36 000 r.p.m., and phas control to $\pm 0{\cdot}3°$. At the Dido reactor at Harwell an exper ment, to study the interaction of resonance-region neutrons wit nuclei, started by using two choppers operating at 12 000 r.p.m in phase to $\pm 0{\cdot}1°$, and a further two rotors operating at ha this speed with phases within $\pm 1°$ of the fast choppers. Othe spectrometers using these units are planned.[7, 8] The extreme low level of noise and vibration in this unit has allowed it to b used for a measurement of the relativistic red shift.[9]

Through many hours of operation the reliability of the co ponents of the spinning head has been established. The wea on the gas bearings is almost negligible, the only trouble bein occasional sooting-up due to impurities in the air supply, ar the oil seal and damper likewise show very little wear. A discussed earlier, the phase stability is satisfactory and apparent can be maintained indefinitely; these instruments have been ru continuously for periods up to 1 000 hours.

(5) ACKNOWLEDGMENTS

We are grateful for the encouragement of Dr. E. Bretsch of A.E.R.E. throughout the course of this work, for the helpf suggestions of Mr. C. S. King of the Rover Co. Ltd. an Mr. C. Robson of Smith's Aircraft Instruments, and for t assistance of Mr. E. E. Maslin of A.E.R.E. (now at A.W.R.E in some of the spinning tests.

(6) REFERENCES

(1) EGELSTAFF, P. A., GAYTHER, D. B., and NICHOLSON, K. I 'The Slow-Neutron Resonance Behaviour of Plutoniu Isotopes', *Journal of Nuclear Energy*, 1958, **6**, p. 303.

(2) SEIDL, F. G. P., HUGHES, D. J., PALEVSKY, H., LEVIN, J. ' KATO, W. Y., and SJÖSTRAND, N. G.: ' "Fast Choppe Time-of-Flight Measurement of Neutron Resonance *Physical Review*, 1954, **95**, p. 476.

(3) ROTERS, H. C.: 'The Hysteresis Motor—Advances whi Permit Economical Fractional Horsepower Rating *Transactions of the American I.E.E.*, 1947, **66**, p. 1419.

(4) SHIRES, O. L.: 'On a Type of Air Lubricated Journal Bearin N.G.T.E. Pyestock, 1949, Report No. R.61.

(5) ROBINSON, C. H.: 'The Static Strength of Pressure-Fed G Journal Bearings—Jet Bearings', A.E.R.E. Harwell, 19 Report No. R/R2642.

(6) FORD, G. W. K., HARRIS, D. M., and PANTALL, D.: 'Princip and Applications of Hydrodynamic-Type Gas Bearing *Proceedings of The Institution of Mechanical Engine* 1957, **171**, p. 93.

(7) DYER, R. F.: 'Neutron Diffraction and Scattering', ed. G. BACON: A.E.R.E. Harwell, 1959, Report No. M/R28 Section 3.

(8) LOWDE, R. D.: 'The Principles of Mechanical Neutr Velocity Selection', *Journal of Nuclear Energy, Part Reactor Science*, 1960, **11**, p. 69.

(9) HAY, H. J., SCHIFFER, J. P., CRANSHAW, T. E., and EGELSTA P. A.: 'Measurement of the Red Shift in an Accelera System Using the Mossbauer Effect in Fe^{57}', *Physi Review Letters*, 1960, **4**, p. 165.

317.79 : 681.142

The Institution of Electrical Engineers
Paper No. 3337 M
Jan. 1961

N ELECTRICAL ANALOGUE FOR HEAT WAVES IN AN EXOTHERMIC MEDIUM

By R. T. ACKROYD, M.Eng., Ph.D., Member, J. HOUSTOUN, B.Eng., Graduate, J. W. LYNN, M.Sc., Ph.D., Associate Member, and E. MANN, B.Sc.

(*The paper was first received* 12*th April, and in revised form* 8*th July*, 1960.)

SUMMARY

The paper describes the design, construction and operation of an trical analogue used for simulating the conditions of propagation eat waves in an exothermic medium.

The equipment is in two parts. The first is a passive *RC* network ich is the usual form of analogue for heat conduction problems. e second consists of a function generator which calculates the stored rgy released as the wave travels across the medium, together with necessary switching apparatus. The function generator is switched idly along the path of the wave. Its output for any point is a ction of the temperature at that point. As the wave reaches a en point, additional energy is released into the network. This difies the conditions at each point to satisfy equations of the form

$$-\nabla^2\phi + \rho\frac{\partial\phi}{\partial t} = f(\phi)$$

The paper also describes the construction and operation of electronic tches capable of scanning the network and adjusting the conditions imes of the order of 5 millisec per node.

(1) INTRODUCTION

Simulators of the resistor-capacitor type have been used for ny years for the solution of problems of transient heat diffusion passive media.[1,2] The more easily solved problems include se for which constant distributed sources of heat are present d those for which there are variable fluxes or temperatures at e boundaries.

When the heat sources are temperature-dependent and are tributed through the medium the classical simulator becomes mplicated and expensive. The simulator usually proposed in s case has, in addition to the *RC* network, a means of supplying riable currents to each node of the network.[3] The strength of e current injected at each node is controlled by a function nerator at that point. This form of simulator therefore, quires at least as many function generators as there are nodes. e cost of such equipment in a complicated problem may well prohibitive and the complexity sets several problems regarding e matching of components.

The paper describes a simulator for heat waves in an exo- ermic medium, with the rate of heat release dependent on nperature. This simulator uses only one function generator, hich is shared among all the nodes. A description is given of e circuits used for a heat problem and an indication is given the use of the machine for more complicated problems of portance in technology.

The sharing of one function generator among all the nodes is hieved by scanning the node potentials in cyclic order and en applying a constant current at each node for one complete anning cycle. The strength of the current is determined by e function generator, from the value of the corresponding node otential when scanned. Although errors are obviously intro- uced by this procedure, they can be kept within reasonable limits by a suitable choice of scanning speed and time scale. Instead of the currents being in accord with the node potentials continuously, they are made to respond in a step-wise manner to a step-wise approximation to the node potentials.

The heat equation considered is of the form

$$k\nabla^2\theta - c\rho\frac{\partial\theta}{\partial t} = f(\theta)$$

The simulator consists of an *RC* network with currents injected at the nodes, as shown in Fig. 1. The switching arrangements are shown in block diagram form in Fig. 2. The nodes of the network are scanned by an electronic switch and the voltage at each node is applied in turn to the function generator. The output of the function generator goes to a second switch, which charges an auxiliary capacitor at each node. The auxiliary capacitor is arranged to feed, via a high resistance, a node current which is analogous to the rate of release of heat in the exothermic medium. The rate of scanning is chosen so that, even up to 100 nodes, the auxiliary capacitors do not discharge to less than 92% of the initial voltage during the scanning cycle. The block diagram shows one node of the *RC* network connected to the scanning and charging switches, which are both controlled from the Dekatron and pulse amplifier unit.

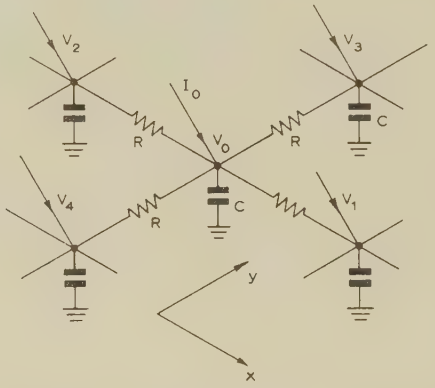


Fig. 1.—A typical node of the *RC* network with currents injected at the nodes.

The output potentials of the scanning switch and the function generator are added together so that the current flowing into the node is proportional to the function-generator output voltage only.

(2) SWITCHING CIRCUITS

(2.1) The Scanning Switches

The scanning switches have to satisfy three conditions:

(*a*) When open, the impedance presented to the network must be very large, i.e. greater than 100 megohms.

(*b*) When closed, the current drawn from the node must be small, i.e. not more than 1μA.

(*c*) The time taken to move from one node to the next must be short, of the order of 1 millisec.

Written contributions on papers published without being read at meetings are vited for consideration with a view to publication.

Dr. Ackroyd and Mr. Mann are with the U.K.A.E.A. Development and Engineering roup, Risley, and Dr. Lynn and Mr. Houstoun are in the Department of Electrical gineering, University of Liverpool.

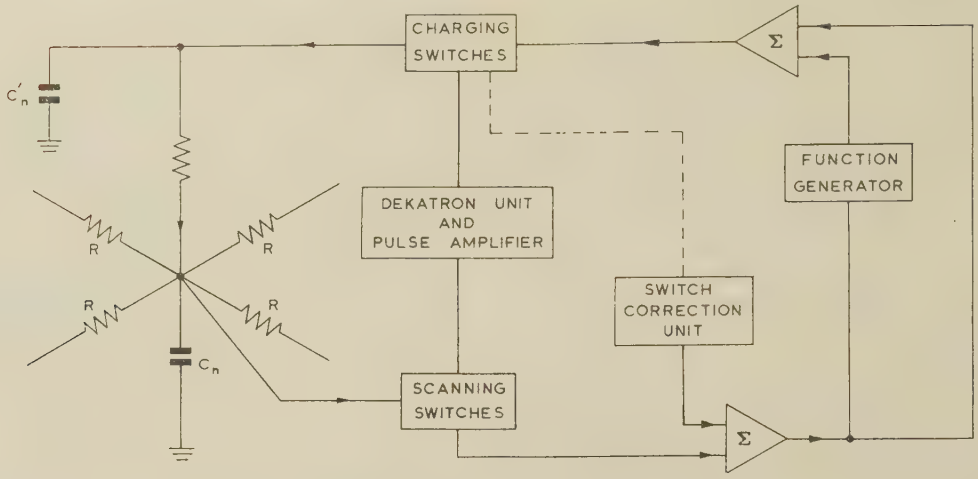


Fig. 2.—Flow diagram of network, switches and function generator.

Condition (*c*) eliminates the use of a mechanical switch such as a uniselector. The alternative is a set of diode switches controlled by a Dekatron.

To meet condition (*a*) it was decided to use silicon diodes with a back impedance greater than 100 megohms. The circuit is shown in Fig. 3. It consists of two diodes connected back to back with the controlling voltage applied to one end, A, of a 6·8-megohm resistor (R_1), the other end being connected to the common point of the two diodes. Normally the control signal is negative and then both diodes are biased off. In this condition the current taken from the node is less than 10^{-8} amp. In the closed condition a positive pulse is applied at A causing both diodes to conduct, and, since their forward resistance is small compared with R_1, the input and output are at the same potential. R_1 limits the current flowing into the switch to 1 μA, so that condition (*b*) is satisfied.

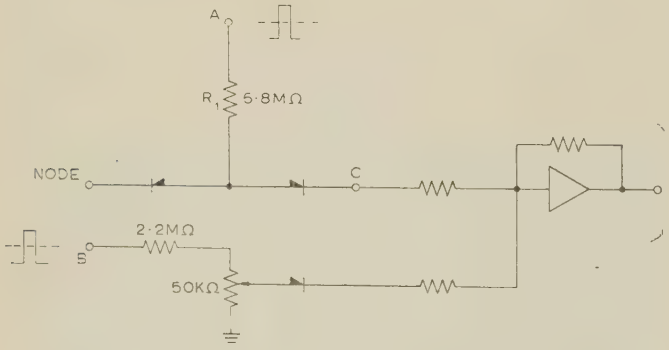


Fig. 3.—A typical scanning switch with its correction circuit.

Since the diodes are being operated near to their cut-off point, there is an appreciable spread in the output current for a constant input, due to the scatter in the diode characteristics. To overcome this source of error, pulses are taken from a point in the charging switch where there is one pulse per node in a scanning cycle, and are applied at B to a potential divider. The output of this is connected through a diode to the input of a summing amplifier and is added to the scanning switch output. This correction circuit is duplicated at each node. By adjusting each voltage divider, all of the switch outputs can be equalized. The outputs of all the switches are connected by a common busbar to the d.c. summing amplifier.

This arrangement will give an output equal to the input, b only when the input is less than one-third of the positive swi of the controlling pulse. However, if another amplifier is co nected back from the output of the first one so that the curre required by the first is provided from the second, then the ran of output of the switch is practically the same as the positi voltage of the pulse. This circuit, shown in Fig. 4, is a standa method of providing a high-impedance probe.[4]

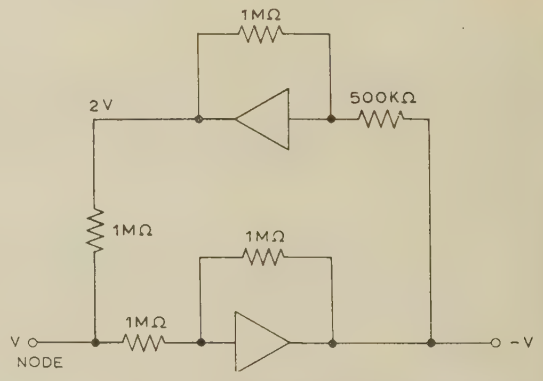


Fig. 4.—A d.c. amplifier arrangement with high input impedance

(2.2) Charging Switches

The charging switches also have to satisfy certain conditio

(*a*) When the switch is open it must not affect the voltage on capacitor C'_n, Figs. 2 and 5. This means that the current flow in the switch output circuit must be less than 10^{-7} amp.

(*b*) When closed, the switch must pass sufficient current to cha or discharge C'_n to the voltage calculated by the function generat

(*c*) The charging and scanning switches must operate synchronism.

To satisfy condition (*b*) the switches must be able to condu in either direction, so an arrangement of four diodes has be used, as shown in Fig. 5. The biasing voltages for these a taken from the anodes of a flip-flop circuit. The positive pu at D is also used for the correction system described in t previous Section. When the input connection from the functi generator is open-circuited at G and the switch is closed, t switch output voltage at F will always be negative. Thus, i decade switching arrangement is used for charging on a lar network, when a capacitor C'_1 is being charged the capaci C'_{11} will be discharged to a negative voltage level unless p cautions are taken. The pulses from the two Dekatrons co

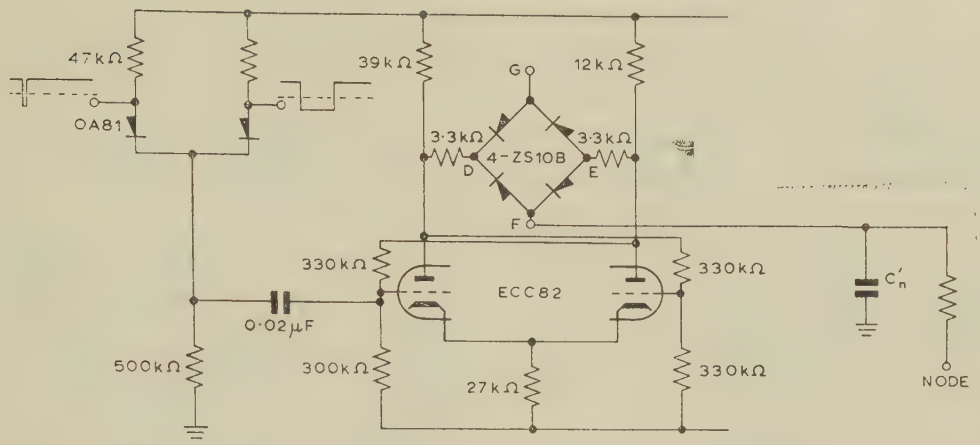


Fig. 5.—The control circuit for a charging switch in a 100-mode network.

ling the switches are therefore applied to the controlling grid he flip-flop through a diode gate, as in Fig. 5, and the switch only close between G and F when both pulses are present. e gating circuit is not necessary for the scanning switches, l G is the most suitable point from which to take pulses for correction circuits.

(2.3) Switch Driving Unit

he switch driving unit consists of a drive oscillator, selector katrons and pulse amplifiers. Originally, a sine-wave oscil- or was used to drive the Dekatrons, but this gave a poor rk/space ratio in the output of the scanning switches and so was changed to a pulse oscillator, with the improvement in io shown in Fig. 6. For a 10-node test network, no pulse plifiers were required for the charging switches, and transistors re used for the scanning switches. However, for a larger work of, say, 100 nodes, the current required to operate the tches would be so much greater that amplifiers would be uired for all pulses.

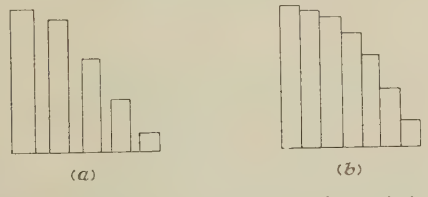


Fig. 6.—The output of the scanning switch.
(*a*) Dekatron driven by a sinusoidal voltage.
(*b*) Dekatron driven by pulses.

(2.4) The Function Generators

The key component of the simulator is the function generator, ich has to calculate the amount of energy released as each int is scanned and signal the appropriate adjustment of node rent. In order to test the switching circuits a straightforward sed-diode function generator was used with 10 diode steps. is gave a close approximation to the square law of heat ease, $S = kT^2$. In a problem of this kind, involving Wigner ergy release, the function would be much more complex and the form $f(T)\partial T/\partial t$. Such a relationship would require a re or a differentiating circuit. Even for simpler functions, biased-diode type may be unsuitable if the function has ferent values for increasing and decreasing temperatures. The general case will therefore require either special analogue nction generators or a digital computing device. The latter quires digital-analogue converting circuits. The proposal to use a digital computer to calculate the current to be injected at each node of the network requires a little explanation, since first thoughts suggest that it would be more convenient and economic to use a digital machine for the complete solution than to share the solution between an analogue and a digital machine.

For the digital solution of a typical problem of heat waves in an exothermic medium[6] a satisfactory technique is briefly as follows. The temperatures, as a function of position at time t, are used to calculate the heat released during the interval t to $t + \Delta t$ as a function of position. The temperatures at time $t + \Delta t$ as a function of position can then be determined by a relaxation process. In this step-by-step calculation, the greater part of the computing time is taken up by the relaxation process, which makes the digital computation slow.

An advantage of the electrical analogue is that this relaxation process is carried out quickly and continuously by the network, instead of a disjointed discrete numerical technique as in the case of the digital machine. On the other hand, an advantage of the digital machine is that it has a very efficient memory and can carry out more conveniently than analogue devices some of the calculations for the scanning unit. Combining a network and a digital machine via a scanning system would appear to provide a faster and more flexible means of solving some problems on heat waves in exothermic media than can be provided by either straight analogue or digital machines.

If a digital computer is used it will be necessary to ensure that the output from the computing circuit is available at the network analogue at a time suitable for the charging switches, say one scanning step after the voltage is measured, and that the charging and scanning switches are properly synchronized.

(3) RESULTS

The apparatus was assembled and the individual components, switches, amplifiers, etc., were tested. A 10-node network was then set up to simulate a block of exothermic material. The following data were used for the system:

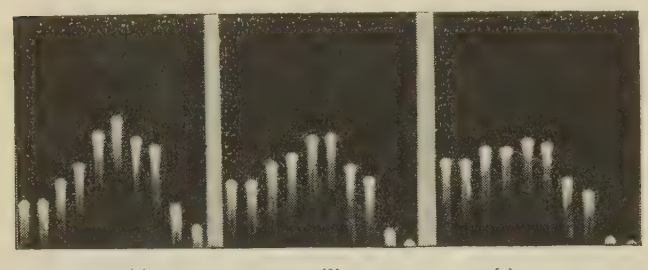
Specific heat	0·3 cal/g per deg C
Density	2·5 g/cm^3
Thermal conductivity	0·33 cal/cm^2/cm per deg C
Node spacing	2 cm
Network resistors	72 kilohms
Network capacitors	2μF
Time scale	1/60 real time
Temperature scale	100° C/volt

The switches were found to operate in the following times:
Scanning: each node scanned 20 times/sec.
Charging time, approximately 4 millisec.
Discharging time to 99% voltage, 50 millisec approximately.

The accuracy of the apparatus obviously depends to a large extent on the relative speeds of the scanning device and the propagation of the wave. In the present case the velocity of propagation was 1·5 nodes/sec, and the above figures show that the circuit was being scanned several times while the wave travelled between nodes.

The behaviour of the medium was recorded under sudden changes of terminal temperatures, with standing and travelling waves. Instability of the thermal state was demonstrated with the square-law heat-release function. Runaway conditions were seen to arise following both step and pulse changes in temperature. In some cases runaway was triggered off, not by the incident wave associated with the pulse travelling across the medium for the first time, but by the superposition of the reflected thermal wave from the far boundary and the tail of the incident wave.

Typical traces are shown in Fig. 7, from which characteristic curves may be derived, such as those in Fig. 8.



(*a*) (*b*) (*c*)

Fig. 7.—Outputs of scanning switch at 0·5 sec intervals during a typical run.

(4) CONCLUSION

The idea of scanning a network analogue has been used in field plotting for many years. Usually this meant a step-by-step analysis programmed for a desk machine or automatic computer. In a recent paper Altes[5] describes the use of mechanical uniselector switches to obtain an iterative solution of the steady-state 2-dimensional field problem described by Poisson's equation, in which $\nabla^2\phi = 4\pi\rho$, where ρ may be a function of ϕ. The automatic scanning and charging circuits described in the paper can be directly and usefully applied to a wide range of such problems, even without the additional complication of time as an independent variable. Iterative solutions can also be obtained for ordinary non-linear equations and for equations with variable coefficients.

Additional problems to which the present simulator could be applied are those of rapid combustion or explosion of gases or liquids, or the build-up of poisoning products in a nuclear reactor. All of these involve multiplicative processes in which energy changes are dependent on temperature, flux, etc., which vary from point to point in space.

(5) ACKNOWLEDGMENTS

The authors wish to thank Dr. Kronberger, Director of Research and Development, United Kingdom Atomic Energy Authority (Development and Engineering Group), Risley, and Professor J. M. Meek, University of Liverpool, for their interest, encouragement and material support. They are also indebted to Dr. G. Black, Technical Manager, Computing Section, Development and Engineering Group, U.K.A.E.A., and Mr. E. Hicks, Safeguards Division, U.K.A.E.A., for discussions on combined analogue and digital machines.

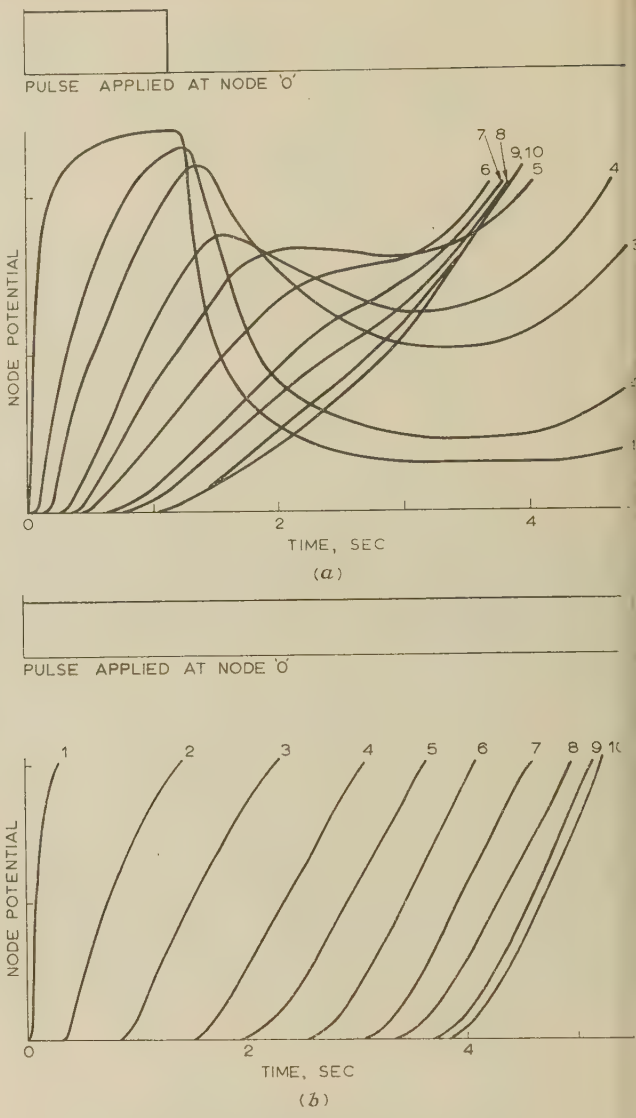


Fig. 8.—Characteristic curves.
(*a*) An unstable solution initiated by a pulse (applied pulse not to scale).
(*b*) An unstable solution initiated by a step change of input.

(6) REFERENCES

(1) PASCHKIS, V., and BAKER, H. D.: 'A Method for Determinii Unsteady State Heat Transfer by means of an Electric Analogy', *Transactions of the American Society of Mec anical Engineers*, 1942, **64**, p. 105.

(2) LAWSON, D. I., and MCGUIRE, J. H.: 'The Solution Transient Heat Flow Problems by Analogous Electri Networks', *Proceedings of The Institution of Mechani Engineers*, 1953, **167** A, p. 275.

(3) ROGERS, T. A.: 'Electronic Analog Computers and Part Differential Equation Solutions', Memorandum of t Department of Engineering, University of California, L Angeles, Nov. 17, 1952.

(4) KARPLUS, W. J.: 'Analogue Simulation' (McGraw-H 1958).

(5) KORTHALS ALTES, J. P.: 'Use of an Electronic Analog Computer with Resistance Network Analogues', *Brit Journal of Applied Physics*, 1959, **10**, p. 176.

(6) HOSKIN, N. E., and COOPER, D. C.: 'Thermal Waves Irradiated Graphite', *Quarterly Journal of Mechanics a Applied Mathematics*, 1959, **12**, p. 393.

.317.312.029: 621.373.42

The Institution of Electrical Engineers
Paper No. 3356 M
Jan. 1961

THE CONCEPT OF EQUIVALENT SOURCE E.M.F. AND EQUIVALENT AVAILABLE POWER IN SIGNAL-GENERATOR CALIBRATION

By D. WOODS, Associate Member.

(*The paper was first received* 9*th April, and in revised form* 11*th August*, 1960.)

SUMMARY

he source impedance of a signal generator, for use under terminated ditions, has a direct effect on the accuracy of the load voltage. nal-generator specifications rarely give detailed information about variation of this parameter with frequency. Consequently, the r is generally forced to assume that it is constant and equal to the ninal value. The view is maintained that it is impracticable to ly corrections for this parameter even if its value is known.

hree methods are analysed for calibrating the output level of signal erators with a view to determining that method which minimizes error in the load voltage due to source impedance variation. The hods are each compared in terms of the voltage ratio produced oss an arbitrary load impedance by an imperfect generator to a fect generator. The voltage ratios are expressed in terms of the ection coefficients of the generator and load impedances. The nimum and maximum of the ratios are given in terms of voltage nding-wave ratios. The results are equally applicable to coaxial l waveguide systems.

he recommended method employs the concept of equivalent source .f. or equivalent available power. The requirements for a suitable rument to measure these quantities are briefly discussed.

Recommendations are made regarding the manner of specifying the rances on signal-generator parameters which affect the accuracy the load voltage, namely the calibration level, the attenuation and source impedance.

LIST OF PRINCIPAL SYMBOLS

E_n = Source e.m.f. of perfect signal generator.
E_s = Source e.m.f. (general case).
E_{sa}, E_{sb}, E_{sc} = Source e.m.f.'s of imperfect signal generator after calibration, methods (*a*), (*b*) and (*c*), respectively.
P_b = Available power of imperfect signal generator before calibration, method (*b*).
P_L = Power dissipated in Z_L (general case).
P_{Lb}, P_{Lc} = Powers dissipated in Z_L by imperfect signal generator, methods (*b*) and (*c*), respectively.
P'_{Lc} = Modified value of P_{Lc} due to imperfect wattmeter.
P_{Ln} = Power dissipated in Z_L by perfect signal generator.
P_n = Available power of perfect signal generator.
S_L = V.S.W.R. of Z_L.
S_s = V.S.W.R. of Z_s.
S_v = V.S.W.R. of voltmeter.
S_w = V.S.W.R. of wattmeter.
V_a = Open-circuit voltage of imperfect signal generator before calibration, method (*a*).
V_c = Voltage across R_n produced by imperfect signal generator before calibration, method (*c*).
V_L = Voltage across Z_L (general case).
V_{La}, V_{Lb}, V_{Lc} = Voltages across Z_L produced by imperfect signal generator, methods (*a*), (*b*) and (*c*), respectively.
V'_{Lc} = Modified value of V_{Lc} due to imperfect voltmeter.
V_{Ln} = Voltage across Z_L produced by perfect signal generator.
Z_L = Load impedance.
Z_n = Nominal characteristic impedance of coaxial system.
Z_s = Source impedance.
ρ_L = Reflection coefficient of Z_L.
ρ_s = Reflection coefficient of Z_s.
ρ_v = Reflection coefficient of voltmeter.
ρ_w = Reflection coefficient of wattmeter.
ϕ_L = Phase angle of ρ_L.
ϕ_s = Phase angle of ρ_s.

(1) INTRODUCTION

The source impedance of a signal generator is usually assumed to be constant with frequency, although in practice considerable variation occurs in this parameter. Source voltage standing-wave ratios (v.s.w.r.) of between 0·5 and 0·7 are typical of signal generators in current use. Departure of the source impedance from the nominal value results in an error in the voltage produced across the load impedance which will be additional to any calibration error in the output level. The error due to source-impedance variation could, of course, be corrected, but this would require a complete knowledge of the impedance/frequency characteristics of both the source and the load. The application of corrections in this manner would be a very tedious procedure and quite unacceptable for routine work. In practice, therefore, one is forced to assume that the source impedance is constant and equal to the nominal value. This raises the question whether there is some optimum way of calibrating the output level so as to minimize the error in load voltage caused by source-impedance variation. This matter was considered in some detail in Reference 1. It was found that the errors could be reduced considerably by adopting the concepts of equivalent source e.m.f. or equivalent available power, which are defined as follows:

Equivalent source e.m.f. is twice the voltage produced across a resistive load equal to the nominal source impedance.

Equivalent available power is the power dissipated in a resistive load equal to the nominal source impedance.

The two concepts are compatible in that a signal generator calibrated by either method will lead to the same estimate of the power dissipated in any load impedance.

In signal-generator calibration, it is normal practice to calibrate the output voltage or power level in two stages. The output is first calibrated at a convenient fixed level, called the calibration level, by means of a voltmeter or wattmeter. Secondly, the attenuator is calibrated in relation to the fixed level by means of a standard attenuator. In this way the output level at all settings of the attenuator is established.

Written contributions on papers published without being read at meetings are ited for consideration with a view to publication.
Mr. Woods is at the Ministry of Aviation.

There are two basic types of signal generator in current use: those having a very low source impedance for use under substantially open-circuit conditions, and those having a source impedance associated with coaxial systems for use under terminated conditions. Signal generators considered in the paper are of the latter type because the only way of calibrating the output level of the former type is in terms of source e.m.f.

(2) WAYS OF ESTABLISHING CALIBRATION LEVEL

There are three common methods for establishing the calibration level, namely:

(*a*) Measurement of the source e.m.f.
(*b*) Measurement of available power.
(*c*) Measurement of the voltage produced across a resistive load equal to the nominal source impedance.

Method (*a*) entails the measurement of the open-circuit voltage by means of a voltmeter having a negligibly high impedance[2] or by measuring the voltage across or current through a number of known resistive loads and thereby deducing the open-circuit voltage.[3] Both techniques are subject to errors of a few per cent and cannot be used at frequencies higher than about 30 Mc/s and 100 Mc/s, respectively.

Method (*b*) demands a wattmeter which will provide a conjugate-impedance match to the generator. The attainment of such a match necessitates the use of tunable stub-matching units for v.h.f. or tunable π-matching units for lower frequencies. These devices, besides being inconvenient, often absorb a significant amount of power owing to their losses. This results in a corresponding error in the measured power.

Method (*c*) is the one which leads to the concepts of equivalent source e.m.f. and equivalent available power. As will be shown subsequently, it is the preferred method because it minimizes the errors due to source-impedance variation by introducing a controlled error in the source e.m.f., or the available power, of such a magnitude that a considerable measure of compensation is obtained. A generator calibrated in this way will, of course, produce no error in the amplitude of the signal in a resistive load equal to the nominal source impedance, because this is the condition of calibration. For a load not equal to the nominal value there will be an error in the signal amplitude the magnitude of which will be dependent on the value of the load and the error in the source impedance. The error in the signal level, however, will be small compared with the departure of the load and source impedances from their nominal values. For example, if both the source and load impedances are 20% different from the nominal value, the error in the voltage across the load will be only about 1% compared with that which would be produced by a perfect signal generator.

(3) COMPARISON OF THE CALIBRATION METHODS

In order to compare the merits of the three methods of measuring the calibration level the concept of a perfect signal generator will be adopted. This is defined as a generator having a frequency-independent source resistance, R_n, equal to the nominal source impedance and a source e.m.f., E_n, equal to the nominal value at the chosen calibration level. The characteristic impedance (resistive) of the coaxial system associated with the generator is also assumed to be equal to R_n.

The various processes involved in calibrating the imperfect signal generator by the three methods are given later together with formulae for the ratio of the voltage produced across a load by the imperfect generator to that by the perfect generator. The formulae are in terms of the reflection coefficients of the source and load impedances.

Before deriving these formulae it is first necessary to establis certain relationships between the parameters of the source, loa and coaxial system.

(4) REFLECTION-COEFFICIENT RELATIONSHIPS

Let the nominal characteristic impedance of the coaxial syste be Z_n; the reflection coefficient ρ of any impedance $Z = 1/$ relative to Z_n, is defined as

$$\rho = \frac{Z - Z_n}{Z + Z_n} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

or

$$\rho = \frac{Y_n - Y}{Y_n + Y} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Rearranging eqn. (1),

$$Z = Z_n \frac{1 + \rho}{1 - \rho} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

ρ is a complex quantity which can be expressed in the followi form:

$$\rho = |\rho| (\cos \phi + j \sin \phi) \quad . \quad . \quad . \quad . \quad (4)$$

If eqn. (4) is substituted in eqn. (3), the denomina rationalized and the expression separated into real and imagina parts, then, provided that Z_n is real and equal to R_n, the re part can be equated to the real part of Z. Thus

$$R = R_n \frac{1 - |\rho|^2}{|1 - \rho|^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Similarly, for the admittance case, eqn. (2),

$$Y = Y_n \frac{1 - \rho}{1 + \rho} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

and

$$G = G_n \frac{1 - |\rho|^2}{|1 + \rho|^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

(5) SOURCE AND LOAD: GENERAL CASE

Consider Fig. 1, which represents a source of e.m.f. E_s a impedance Z_s connected to a load of impedance Z_L by mea

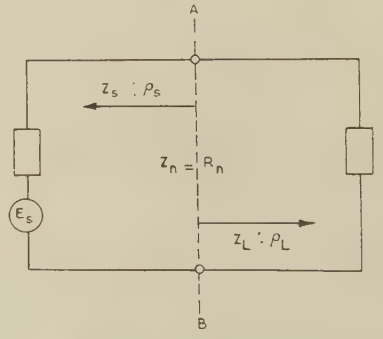


Fig. 1.—Source and load.

of coaxial connectors, of characteristic resistance R_n, the junctio of which is in the plane AB. The reflection coefficients in th plane are ρ_s and ρ_L. The connectors form part of the sour and load.

Let the voltage produced across Z_L, in the plane AB, be V Then

$$V_L = E_s \frac{Z_L}{Z_s + Z_L}$$

xpressing Z_s and Z_L in terms of Z_n and their reflection fficients in accordance with eqn. (3),

$$V_L = \frac{E_s}{2}\frac{(1-\rho_s)(1+\rho_L)}{1-\rho_s\rho_L} \quad . \quad . \quad . \quad . \quad (8)$$

s is the general case for the voltage across the load in terms he source e.m.f. and the reflection coefficients of the source load, and is a complex quantity. The general case for the er, P_L, absorbed by the load is given by the modulus squared V_L times the conductance G_L of the load. The conductance erms of the reflection coefficient is given by eqn. (7). refore

$$P_L = \frac{|E_s|^2}{4R_n}\frac{|1-\rho_s|^2(1-|\rho_L|^2)}{|1-\rho_s\rho_L|^2} \quad . \quad . \quad . \quad . \quad (9)$$

is important to realize that eqns. (8) and (9) are valid only the reflection coefficients at the plane of cleavage, AB. Any xial cable attached to the source or load must be included in measurement of ρ_s or ρ_L. In other words, the interposition a length of coaxial cable between the connectors at which nd ρ_L are determined will invalidate the above relationships ss either ρ_s or ρ_L (or both) is zero and the characteristic edance of the cable is equal to R_n.

(6) SOURCE E.M.F. METHOD (*a*)

ig. 2(i) represents the perfect signal generator. When it is d with a load impedance, Z_L, the voltage across it will be, ccordance with eqn. (8) with $\rho_s = 0$,

$$V_{Ln} = \frac{E_n}{2}(1+\rho_L) \quad . \quad . \quad . \quad . \quad . \quad (10)$$

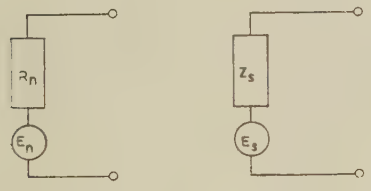


Fig. 2.—Equivalent circuit of signal generator.
(i) Perfect.
(ii) Imperfect (before calibration).

ig. 2(ii) represents the imperfect signal generator before bration. The measured open-circuit voltage, V_a, will be [Fig. 3(i)]. A calibration correction will be applied equal E_n/E_s in order to make the corrected source e.m.f., E_{sa}, equal E_n. When the generator is subsequently used with a load edance Z_L [Fig. 3(ii)] the voltage across it will be, in accor- ice with eqn. (8),

$$V_{La} = \frac{E_n}{2}\frac{(1-\rho_s)(1+\rho_L)}{1-\rho_s\rho_L} \quad . \quad . \quad . \quad . \quad (11)$$

erefore the ratio of the moduli of V_{La} and V_{Ln} is

$$\left|\frac{V_{La}}{V_{Ln}}\right| = \frac{|1-\rho_s|}{|1-\rho_s\rho_L|} \quad . \quad . \quad . \quad . \quad . \quad (12)$$

s not possible to plot eqn. (12) because ρ_s and ρ_L are complex antities, i.e. there are four variables. However, it is possible obtain the minimum and maximum of the expression, in ms of the v.s.w.r.'s of the source and load, by expressing ρ_s d ρ_L in terms of their moduli and phase angles and then

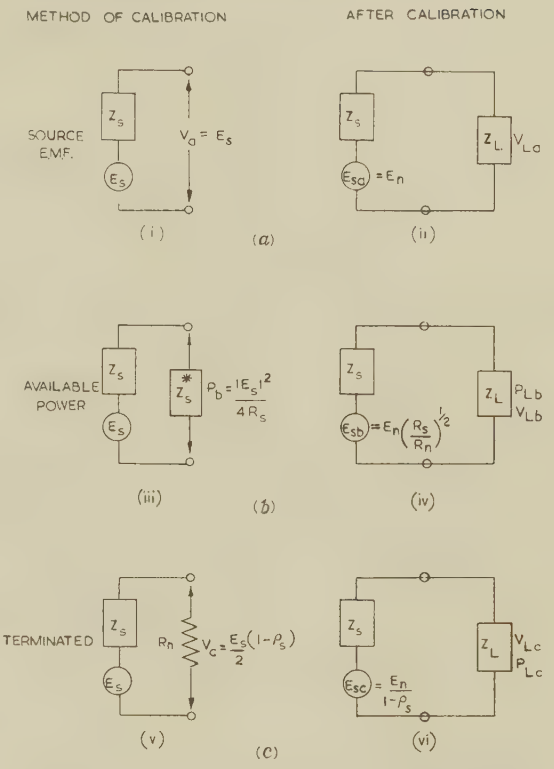


Fig. 3.—Imperfect signal generator (after calibration).

assigning to the angles values which give the worst conditions. Thus, using eqn. (4),

$$\frac{|1-\rho_s|}{|1-\rho_s\rho_L|} = \frac{(1+|\rho_s|^2-2|\rho_s|\cos\phi_s)^{1/2}}{[1+|\rho_s|^2|\rho_L|^2-2|\rho_s||\rho_L|\cos(\phi_s+\phi_L)]^{1/2}} \quad . \quad (13)$$

It will be seen that when the cosines are either $+1$ or -1 the expressions in the brackets are perfect squares. The minimum of eqn. (13) occurs when $\cos\phi_s = 1$ and $\cos(\phi_s+\phi_L) = -1$, and the maximum when $\cos\phi_s = -1$ and $\cos(\phi_s+\phi_L) = 1$. The minimum and maximum of eqn. (13) are therefore

$$\frac{1-|\rho_s|}{1+|\rho_s||\rho_L|} \quad \text{and} \quad \frac{1+|\rho_s|}{1-|\rho_s||\rho_L|}$$

Now $$|\rho| = \frac{1-S}{1+S}$$

where S is the v.s.w.r. (less than unity).

Therefore, the minimum and maximum in terms of the v.s.w.r.'s are

$$\frac{S_s(1+S_L)}{1+S_sS_L} \quad \text{and} \quad \frac{1+S_L}{S_s+S_L} \quad . \quad . \quad . \quad . \quad (14)$$

These expressions are plotted in Fig. 4 [full curves, (*a*)] for $S_s = 0{\cdot}5$ and S_L between 0 and 1. For any given value of S_L the magnitude of $|V_{La}/V_{Ln}|$ can lie anywhere between the full curves, (*a*). The actual value will depend on the phase angles of the two reflection coefficients.

(7) AVAILABLE-POWER METHOD (*b*)

The available power, P_n, of the perfect generator [Fig. 2(i)] will be

$$P_n = \frac{|E_n|^2}{4R_n}$$

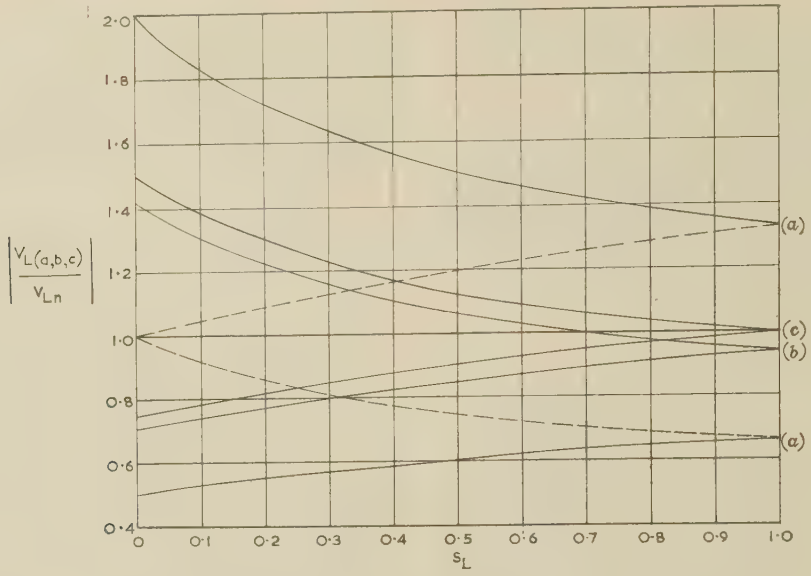


Fig. 4.—Minimum and maximum values of load voltage ratio.

$S_s = 0{\cdot}5$

(*a*) $|V_{La}/V_{Ln}|$ Source e.m.f. method (*a*).
(*b*) $|V_{Lb}/V_{Ln}|$ Available power method (*b*).
(*c*) $|V_{Lc}/V_{Ln}|$ Terminated method (*c*).

The available power, P_b, of the imperfect generator [Fig. 3(iii)] of source impedance Z_s will be delivered when the load impedance is Z_s^*, the conjugate of Z_s. Therefore

$$P_b = \frac{|E_s|^2}{4R_s}$$

In the calibration procedure, E_s will be corrected to E_{sb} in order to make $P_b = P_n$.

Therefore $$E_{sb} = E_n\left(\frac{R_s}{R_n}\right)^{1/2}$$

When the generator is subsequently used with a load Z_L [Fig. 3(iv)] the power absorbed by it will be, in accordance with eqn. (9),

$$P_{Lb} = \frac{|E_n|^2}{4R_n}\,\frac{R_s}{R_n}\,\frac{|1-\rho_s|^2(1-|\rho_L|^2)}{|1-\rho_s\rho_L|^2}$$

Expressing R_s in terms of R_n and ρ_s in accordance with eqn. (5),

$$P_{Lb} = \frac{|E_n|^2}{4R_n}\,\frac{(1-|\rho_s|^2)(1-|\rho_L|^2)}{|1-\rho_s\rho_L|^2} \quad . \quad . \quad . \quad (15)$$

The power delivered to Z_L by the perfect generator will be, in accordance with eqn. (9) with $\rho_s = 0$,

$$P_{Ln} = \frac{|E_n|^2}{4R_n}(1-|\rho_L|^2) \quad . \quad . \quad . \quad . \quad (16)$$

Therefore $$\frac{P_{Lb}}{P_{Ln}} = \frac{1-|\rho_s|^2}{|1-\rho_s\rho_L|^2} \quad . \quad . \quad . \quad . \quad . \quad (17)$$

or, in terms of the modulus of the voltage ratio across Z_L,

$$\left|\frac{V_{Lb}}{V_{Ln}}\right| = \frac{(1-|\rho_s|^2)^{1/2}}{|1-\rho_s\rho_L|} \quad . \quad . \quad . \quad . \quad . \quad (18)$$

Adopting the same procedure as before, the minimum ar maximum values of eqn. (17) in terms of the v.s.w.r.'s are

$$\frac{S_s(1+S_L)^2}{(1+S_sS_L)^2} \quad \text{and} \quad \frac{S_s(1+S_L)^2}{(S_s+S_L)^2} \quad . \quad . \quad . \quad (1$$

Expressions (19) are, of course, power ratios. The minimu and maximum voltage ratios arising from eqn. (18) are therefo

$$\frac{\sqrt{S_s}(1+S_L)}{1+S_sS_L} \quad \text{and} \quad \frac{\sqrt{S_s}(1+S_L)}{S_s+S_L} \quad . \quad . \quad . \quad (2$$

Expressions (20) are plotted in Fig. 4(*b*) for $S_s = 0{\cdot}5$ for com parison with the two other methods which are also in terms voltage.

(8) TERMINATED METHOD (*c*): EQUIVALENT SOURCE E.M.F. OR EQUIVALENT AVAILABLE POWER

In method (*c*) the equivalent source e.m.f. is taken as twi the voltage produced across a resistive load equal to R_n. accordance with eqn. (8) when $\rho_L = 0$, the voltage produced the imperfect generator across R_n will be [Fig. 3(v)]

$$V_c = \frac{E_s}{2}(1-\rho_s)$$

The equivalent source e.m.f. is therefore

$$2V_c = E_s(1-\rho_s)$$

Since the nominal source e.m.f. at the calibration level is E_n, calibration correction, equal to $E_n/E_s(1-\rho_s)$, will be appli in order to make the equivalent source e.m.f. equal to E_n. Th will change the true source e.m.f. from E_s to E_{sc} when

$$E_{sc} = \frac{E_n}{1-\rho_s}$$

When the generator is subsequently used with a load

g. 3(vi)] the voltage across it will be, in accordance with
i. (8),

$$V_{Lc} = \frac{E_n}{2} \frac{1 + \rho_L}{(1 - \rho_s\rho_L)} \quad . \quad . \quad . \quad . \quad (21)$$

om eqns. (21) and (10) we therefore obtain

$$\left|\frac{V_{Lc}}{V_{Ln}}\right| = \frac{1}{|1 - \rho_s\rho_L|} \quad . \quad . \quad . \quad . \quad . \quad (22)$$

opting the same procedure as before, the minimum and
ximum values of eqn. (22) in terms of the v.s.w.r.'s are

$$\frac{(1 + S_s)(1 + S_L)}{2(1 + S_sS_L)} \quad \text{and} \quad \frac{(1 + S_s)(1 + S_L)}{2(S_s + S_L)} \quad . \quad (23)$$

pressions (23) are shown plotted in Fig. 4(*c*) for $S_s = 0{\cdot}5$.

(9) DISCUSSION OF RESULTS

The curves in Fig. 4 are for a source v.s.w.r. of 0·5. For other
ues of S_s approaching unity the pairs of curves (*a*) and (*c*) will
se up and eventually coalesce with the $|V_L/V_{Ln}| = 1$ axis when
$= 1$. For curves (*b*), in addition to closing up, the meeting-
nt on the $S_L = 1$ axis will move up to the $|V_L/V_{Ln}| = 1$ axis.
will be seen that for method (*c*) the error is always zero
$_L/V_{Ln}| = 1$) for $S_L = 1$ for any value of S_s. This is not the
e for the two other methods, where the maximum error is
% for method (*a*) and 6% for method (*b*) when $S_s = 0{\cdot}5$.
t has already been explained that the pairs of curves represent
minimum and maximum of $|V_L/V_{Ln}|$ under the worst con-
ions of phase angles of ρ_s and ρ_L. For other phase angles
error will be smaller. In fact, under a favourable combina-
n of phase angles, the error can be zero for methods (*a*) and
for any value of S_s and S_L. This is not the case for method (*b*)
ause both curves lie below the $|V_L/V_{Ln}| = 1$ axis for a
ited range of S_L below unity. This range decreases as S_s
proaches unity because the meeting-point of the curves at
$= 1$ moves upwards.
t should be remembered that there are always two values of
pedance for a given v.s.w.r., one smaller and one larger than
For methods (*b*) and (*c*) either condition for Z_s or Z_L
ults in the minimum and maximum values given by the curves
Fig. 4. This is not the case for method (*a*), where the full
ves relate to $|Z_L| < R_n$. When $|Z_L| > R_n$ the error is
ays less. The reason for this can be appreciated qualitatively
en it is remembered that the source is calibrated in terms of
n-circuit voltage (source e.m.f.) so that, as $Z_L \to \infty$, V_L will
substantially independent of Z_s, i.e. $|V_L/V_{Ln}| \to 1$. The
its for this condition are given by interchanging the numera-
s and denominators of expressions (14), which are plotted
Fig. 4 by the dotted curves (*a*).

(10) THE CALIBRATING INSTRUMENT

An instrument to measure equivalent source e.m.f. or equiva-
t available power must present a good match to the nominal
aracteristic impedance of the system. For this reason it is
ssible to combine the dual function of voltage and power
asurement in the same instrument. If the instrument
ployed to measure equivalent available power is essentially
oltage-measuring device, in which the power is given by the
oduct of the voltage squared and the nominal conductance,
y departure of the conductance from the nominal value, G_n,
l give rise to an error in determining the calibration level. If
the reflection coefficient of the instrument, treated as a voltmeter, is ρ_v, it can be shown that eqn. (21) is modified to

$$V'_{Lc} = \frac{E_n}{2} \frac{(1 - \rho_s\rho_v)(1 + \rho_L)}{(1 - \rho_s\rho_L)(1 + \rho_v)} \quad . \quad . \quad . \quad (24)$$

The fractional error in the signal-generator calibration level due to ρ_v is therefore given by

$$\left|\frac{V'_{Lc}}{V_{Lc}}\right| = \frac{|1 - \rho_s\rho_v|}{|1 + \rho_v|} \quad . \quad . \quad . \quad . \quad . \quad (25)$$

and the minimum and maximum in terms of the v.s.w.r.'s are

$$\frac{S_s + S_v}{1 + S_s} \quad \text{and} \quad \frac{1 + S_sS_v}{S_v(1 + S_s)} \quad . \quad . \quad . \quad . \quad (26)$$

On the other hand, if the instrument measures power directly in terms of heat produced in the terminating resistor, the error will be considerably less because the error in the voltage will be substantially compensated by the error in the terminating resistor for v.s.w.r.'s near unity. If the reflection coefficient of the wattmeter is ρ_w, a signal generator calibrated by it in terms of equivalent available power will deliver to a load, Z_L, the following power:

$$P'_{Lc} = \frac{|E_n|^2}{4R_n} \frac{1 - |\rho_L|^2}{|1 - \rho_s\rho_L|^2} \frac{|1 - \rho_s\rho_w|^2}{1 - |\rho_w|^2} \quad . \quad . \quad (27)$$

If the wattmeter is perfectly matched, i.e. $\rho_w = 0$, then

$$P_{Lc} = \frac{|E_n|^2}{4R_n} \frac{1 - |\rho_L|^2}{|1 - \rho_s\rho_L|^2} \quad . \quad . \quad . \quad . \quad (28)$$

so that the fractional error in the signal-generator calibration is

$$\frac{P'_{Lc}}{P_{Lc}} = \frac{|1 - \rho_s\rho_w|^2}{1 - |\rho_w|^2} \quad . \quad . \quad . \quad . \quad . \quad (29)$$

and the minimum and maximum in terms of the v.s.w.r.'s are

$$\frac{(S_s + S_w)^2}{S_w(1 + S_s)^2} \quad \text{and} \quad \frac{(1 + S_sS_w)^2}{S_w(1 + S_s)^2} \quad . \quad . \quad . \quad (30)$$

or in terms of voltage ratios

$$\frac{S_s + S_w}{\sqrt{S_w}(1 + S_s)} \quad \text{and} \quad \frac{1 + S_sS_w}{\sqrt{S_w}(1 + S_s)} \quad . \quad . \quad . \quad (31)$$

Assuming a typical value for $S_v = S_w = 0{\cdot}99$ the maximum errors given by expressions (26) and (31) for three values of S_s are illustrated in Table 1.

Table 1

ERRORS IN SIGNAL-GENERATOR VOLTAGE CALIBRATION LEVEL

S_s	Error	
	voltmeter eqn. (26)	wattmeter eqn. (31)
	%	%
1·0	±0·5	+0·0013
0·5	±0·63	±0·18
0	±1·0	+0·5

The errors for the voltmeter case in Table 1 apply to $|Z_s| > R_n$. When $|Z_s| < R_n$ the errors decrease from ±0·5% at $S_s = 1$ to zero at $S_s = 0$. The reason for this is explained in Section 9.

In the case of the wattmeter the errors are the same for both values of Z_s for each value of S_s.

From the foregoing it will be apparent that the matching requirements are less stringent for a calibrating instrument which operates on the principle of heat generated in its load. Because the power at the calibration level is low—about 200 μW—the technique is usually restricted to a thermistor or bolometer device. In addition, the decoupling arrangements for the d.c. power usually limit the lower frequency to about 100 Mc/s. For still lower frequencies it is more convenient to use a voltage-measuring device employing a silicon diode. This enables satisfactory decoupling to be achieved down to the lowest radio frequencies. The matching arrangements, however, call for extreme care in design, and provision must be made for compensation of the diode admittance.[4]

(11) CONCLUSIONS

From the foregoing analysis it is clear that the best way of calibrating a signal generator is in terms of equivalent source e.m.f. or equivalent available power. These concepts are necessary only when the source impedance is not equal to the nominal value and it is impracticable to apply corrections. Once a generator has been so calibrated it is not necessary to retain either of these concepts. The generator can then be considered as having a true source e.m.f. or true available power, as established by this method, and a constant source impedance equal to the nominal value. It will be apparent that the equivalent source e.m.f. or equivalent available power will have to be established at a sufficient number of frequencies to take into account the degree of source-impedance variation with frequency. Signal generators operating at the higher frequencies exhibit a fairly rapid change in source impedance with frequency, especially when used with a length of coaxial cable. The longer the cable the more rapid the change in impedance will be for the same frequency change. For example, if the cable is one wavelength long then for a 10% change in frequency the source impedance, as seen from the remote end of the cable, will traverse approximately one-fifth of the impedance circle. For a cable ten wavelengths long the impedance will describe two revolutions for the same frequency change.

It has already been explained that the output level of a signal generator is usually established in two stages, namely measurement of the calibration level and calibration of the attenuator system. In a typical signal-generator specification a tolerance will be assigned to both these parameters. If the calibration level is established in terms of the equivalent source e.m.f. or equivalent available power, an additional uncertainty will hav to be assigned to the output level to allow for the effect o source-impedance variation. This uncertainty can be co trolled only by specifying a tolerance on the source impedanc It is clear, therefore, that the tolerances on these three paramete (calibration level, attenuation and source impedance) are inte related and that they should bear the same relation to one anoth whatever the grade of the generator. In Reference 1 it has bee recommended that each tolerance should be such that, under th worst conditions, each contributes equally to the total unce tainty on the output level. For example, consider a generat of such an accuracy that the voltage produced across a load any value is not to be in error by more than 3% compared wi a perfect generator, then 1% would be allowed on the calibr tion level, 0·1 dB on the attenuator, and the v.s.w.r. of the sour would have to be not less than 0·98. In the latter respe expressions (23) give a minimum and maximum of 0·99 a 1·01 for $S_s = 0{\cdot}98$ and $S_L = 0$.

It is emphasized that the impedance/frequency characteris of the source need be verified only once, i.e. it can be treated a type test. Having ascertained that it is within the permiss tolerance over the complete frequency range of the generat it can then be ignored provided that the calibration leve determined by method (*c*) at the requisite number of frequenc

Although the analysis has been carried out with coax systems in mind it is, of course, equally applicable to waveg systems, because the results are presented in terms of reflec coefficients or v.s.w.r.'s. The concept of source and load volta is usually meaningless in waveguide systems and it is customa to adopt the power concept. In this instance the expressio for minimum and maximum voltage ratios in terms of v.s.w can be simply converted to power ratios by squaring.

(12) REFERENCES

(1) Radio and Electronic Measurements Committee: 'Jo Service Standards and Recommendations for Sig Generator Calibration', Report No. REMC/24/FR, Issue December, 1960.

(2) Stride, P. L., and Platt, J. D.: 'A Crystal Millivoltmeter f Use between 1 Mc/s and 300 Mc/s', Unpublished Minist of Supply Report No. A.I.D./Rad. Elect./3; August, 19

(3) Coulson, N.: 'A Thermal Millivoltmeter for Measuri Radio-Frequency Voltages', *Proceedings I.E.E.*, Pap No. 1004 R, September, 1950 (**97**, Part III, p. 344).

(4) Woods, D.: 'A Coaxial Millivoltmeter/Milliwattmeter f Frequencies up to 1 Gc/s' (to be published).

.142:662.6.002.71

The Institution of Electrical Engineers
Paper No. 3324 M
Jan. 1961

AN ELECTRONIC ANALOGUE COMPUTER FOR A COAL TRANSPORTATION PROBLEM

By E. G. ANDERSON, B.Sc., A.Inst.P., Associate Member.

(*The paper was first received 8th February, and in revised form 21st June*, 1960.)

SUMMARY

he transportation problem and linear programming are defined and electronic analogue of the problem and its optimum solution is cribed.

A computer with a 5 × 20 matrix has been built using current and posing e.m.f. to simulate coal flow and route costs respectively. e circuits are described and a Section on operating technique is luded.

A typical problem and its solution are given, and it is concluded t a larger computer is feasible.

(1) INTRODUCTION

The transportation problem is to move goods from supply nts to consumers at the minimum cost. Factors other than nsport costs are often involved; for instance, one supply point y need to have its products distributed between several con- ners irrespective of transport cost (an example of this is coal h high chloride content being spread between several power tions), or production capacity may be greater than required that storage costs must also be considered. Problems such these are in the class of linear programming, which may be ined as follows: if a set of simultaneous equations has an inite number of solutions, linear programming is the deter- nation of the solution which optimizes the value of a certain nction of the variables.

The transportation problem is well known,[1] and techniques ve been proposed for solving this or similar problems by hand mputation[2] and by digital[3] and general-purpose analogue[4] mputers. The paper describes a special-purpose analogue mputer built to aid the correct solution of the allocation of il to C.E.G.B. power stations. The computer is, however, plicable to any problems of a similar nature, e.g. coal from nes to gasification plants or packets of detergent from factories wholesale depots.

The coal-transportation problem of the Central Electricity nerating Board had already been described;[5] it involves the cation of more than 40 million tons of coal per year from) mines to 200 power stations, with a total transport cost of ne £30 million. Various limitations exist; for example, route pacity and coal suitability both exert restrictions on the blem, which is not static, since the coal requirements of wer stations can vary widely from week to week, as can the il-mine outputs.

With these difficulties the problem is not at present soluble, t improvements to the coal flow can be brought about by timizations of smaller problems, the sizes of which are defined geographical or administrative factors. As an aid to this o-optimization process and as a guide to the feasibility of a ssible future larger computer, an analogue computer has been ilt at the Central Electricity Research Laboratories. This is a cial-purpose computer which automatically attains a state equivalent to a minimum-cost solution. It uses current to represent coal flow and opposing e.m.f.'s to simulate route costs —a system suggested by Crossman.[6]

Problems which can be set on the computer are limited to those where the sum of the pit outputs is equal to the sum of the power-station inputs; unbalanced problems can be solved by providing an extra power station or coal mine with zero route costs (known as slack variables) to take the surplus.

(2) THE ANALOGUE

Consider the electrical network shown in Fig. 1(*a*). It consists of n current sources, A_j, and m current sinks, B_i, respectively feeding current into and taking current from the matrix. V_{ij} are batteries which oppose the current flow, I_{ij}, between

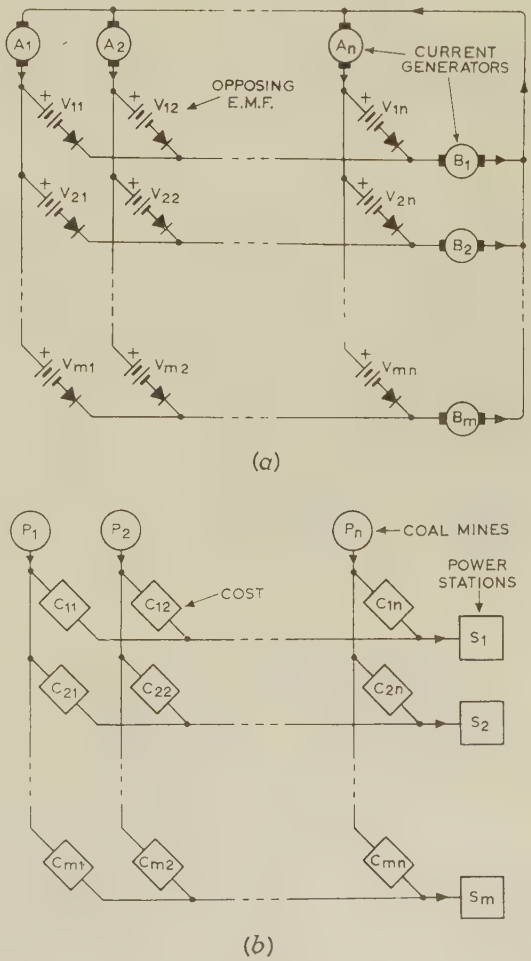


Fig. 1.—The analogous networks.
(*a*) Electrical network
(*b*) Transport network.

Written contributions on papers published without being read at meetings are ted for consideration with a view to publication.

Mr. Anderson is with I.B.M. United Kingdom, Ltd., and was formerly at the Central ctricity Research Laboratories.

generators B_i and A_j; the diodes prevent reverse current flows and ensure that the batteries do not supply power to the matrix.

The equations governing the values of I_{ij} are

$$\left.\begin{aligned} \sum_{i=1}^{m} I_{ij} &= A_j \\ \sum_{j=1}^{n} I_{ij} &= B_i \\ I_{ij} &\geqslant 0 \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

It can be shown[7] that, in any system of steady currents, the currents are distributed so that the heat generated (or power dissipated) is a minimum. If the diodes are perfect, the power dissipated in the matrix is the sum of the products of the battery e.m.f.'s and the currents flowing through them; therefore

$$\sum_{i=1}^{m} \sum_{j=1}^{n} I_{ij}V_{ij} \text{ is a minimum} \quad . \quad . \quad . \quad . \quad (2)$$

The transport network shown in Fig. 1(*b*) shows n coal mines supplying m power stations via routes with costs C_{ij}. The coal distribution is evidently governed by equations similar to (1). If A_j, B_i and V_{ij} are proportional to P_j, S_i and C_{ij} respectively, the coal-distribution equivalent to the steady state of network A is

$$\sum_{i=1}^{m} \sum_{j=1}^{n} X_{ij}C_{ij} \text{ is a minimum} \quad . \quad . \quad . \quad . \quad (3)$$

This condition is that of minimum total transport cost and so the distribution of current in the electrical network is equivalent to a coal distribution in the transport network with minimum transport cost.

(3) GENERAL DESCRIPTION

Fig. 2 shows a schematic of the computer in abbreviated form with two power stations, S_1 and S_2, and three pits, P_1 P_2 and P_3. Since the circuit for a current source is more complicated than that for a current sink and there are more pits than power stations, pits are represented by the simpler current sinks; in other words, negative current represents coal flow. C_{ij} are constant-e.m.f. units to simulate route costs and are adjusted by means of the potentiometers shown.

In operation, the current generators are set to their required values with the switches, S, closed so that the generator outputs are short-circuited and no current flows in the matrix. If the required output of a power-station current source is greater than the capacity of a single generator, two or more sources can be paralleled, as in S_2. The cost units, C_{ij}, are also set to their

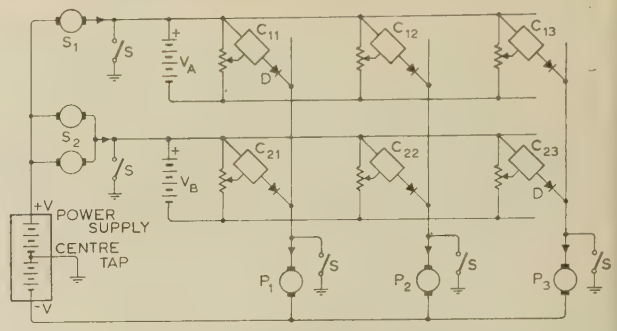


Fig. 2.—Schematic of computer.

appropriate values and switches S are then opened. Under these conditions all current from the current sources passes through the matrix and then out via the current sinks. The system is then monitored to discover conducting paths, and with these results the original coal-transportation problem can be solved directly.

The computer has a matrix size of 5 × 20, i.e. five power stations and 20 pits, but there are ten current sources to provide the paralleling facilities mentioned above. There are two milliammeters, one for setting the current generators and the other for measuring current flow in the constant-e.m.f. units. Two voltmeters are fitted; one measures the voltage across a cost unit and the other the voltage across the cost unit and diode, i.e. the voltage existing between current-source output and current-sink input.

(4) OPERATION

(4.1) Scaling

Before being set into the computer, coefficients must be adjusted to lie in the following ranges:

Cost	3(0·5)23 volts
Pit output	0(0·1)10mA
Power station input	0(0·1)50mA

with the restriction that the sum of pit outputs must not be more than 100mA.

When obtaining route costs it is desirable that the percentage differences be as large as possible, so all base costs (i.e. row + column constants) should be subtracted before scaling takes place. For example, in Table 1, 18 is subtracted from column 1, 139 from column 2 and so on.

Table 1

ACTUAL PROBLEM WITH ROUTE COSTS (IN PENCE PER TON), PIT OUTPUTS AND POWER-STATION INPUTS

Input to power stations																					
1	342	192	208		281	128	287	82	177		56	38	239	195	163			155	129	251	
2	35	62	139		205	144	47		20	175	101		124	118		217	177	182	180	102	
3	140				92	101			32	182	14	26			124	165	145	206	203	119	111
4	77	49	152	114	148		133	98		256	109	155	221	151	158	208		218	216		140
5	186	18		89	247		69	8		220	60	84		134		193		86	84		183
100's of tons per week →		49	17	21	54	3	96	17	38	61	32	73	44	21	27	79	12	18	23	30	65
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

Pit outputs

(4.2) Setting Up

When the values have been arranged to lie in the ranges mened above they are set into the machine; where no route cost available the cost unit is removed or set at maximum. If any wer station requires more than 10 mA, one or more of stations 10 are paralleled to provide the extra current.

At this stage all the earthing switches except one are opened. e generator associated with this switch still has its output ort-circuited, but all the other generators pass current through matrix. The remaining earth link carries a current equal to algebraic sum of the current sinks and sources, since the sum rents into and out of the matrix must be equal. If all erators have been set correctly, the measured current in the k should equal the set value of the remaining generator. This a powerful check on the correctness of setting up the current erators. When the last earth switch is opened the network is e to attain the state of minimum energy dissipation and the rrect solution can be found.

(4.3) Finding the Solution

The solution can be found in two ways. The first method uses e two voltmeters: if the voltage across the cost unit and diode the same as that across the unit alone, the diode is conducting d that route is in use; if it is less, the diode is biased off and e route is not in use. In the latter case the 'unit plus diode' ltage represents the break-even cost of route, i.e. the maximum st of transport at which use of that route would be economic, d this information can be of use. The second method is to easure the current in each route; this has the advantage that, ot only are conducting routes found, but also the amount of nduction.

Probably the best method is the combination of both which oduces a complete solution of the problem giving both actual stribution and break-even costs on unused routes.

Information on the quality of conducting routes is also ailable. If the cost of a conducting route is raised, it is easy see at what cost it becomes uneconomic. One of the main lvantages of an analogue computer—its realism—is particularly oticeable in this case, since the effect of altering costs can be bserved directly.

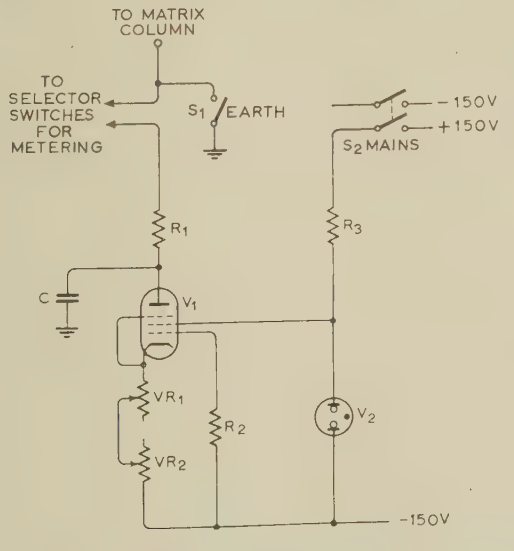


Fig. 3.—Circuit of current sink.

V_1, CV4010 R_1, 100Ω
V_2, CV4048 R_2, 100kΩ
VR_1, 25kΩ R_3, 27kΩ
VR_2, 250Ω C, 0·1μF

(5) CIRCUIT DESCRIPTION

For the current generators (sinks and sources) a pentode, CV4010, is used as a constant-current supply. The current sink as shown in Fig. 3 uses a neon stabilizer, CV4048, to keep the screen at a constant potential above the cathode; VR_1 and VR_2 are coarse and fine current controls used to vary the grid-cathode voltage. R_1 and C are to prevent spurious oscillations.

The circuit of the current source (Fig. 4) is similar but more complicated, because the cathode of V_1 has a variable potential. This is overcome by using a second pentode as a cathode-follower to establish a low-impedance voltage source equal to that of the cathode. The two cathodes of the stabilizer and current source can then be maintained equal without altering the current flow through V_1.

The circuit of a constant-e.m.f. unit, shown in Fig. 5, is based

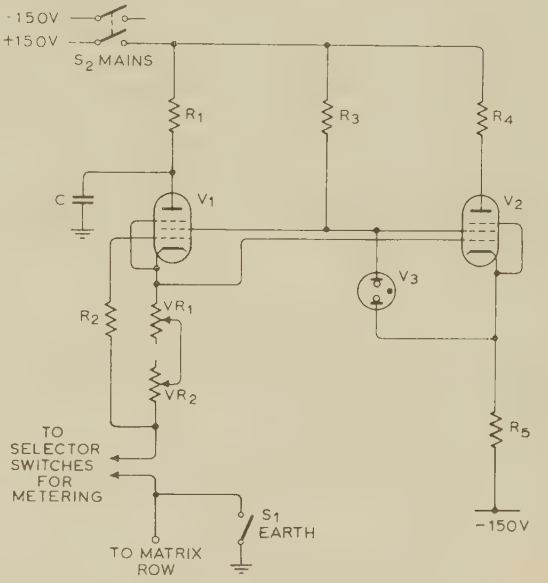


Fig. 4.—Circuit of current source.

V_1, CV4010 VR_2, 250Ω R_3, 6·8kΩ
V_2, CV4010 R_4, 100Ω
V_3, CV4048 R_1, 100Ω R_5, 10kΩ
VR_1, 25kΩ R_2, 100kΩ C, 0·1μF

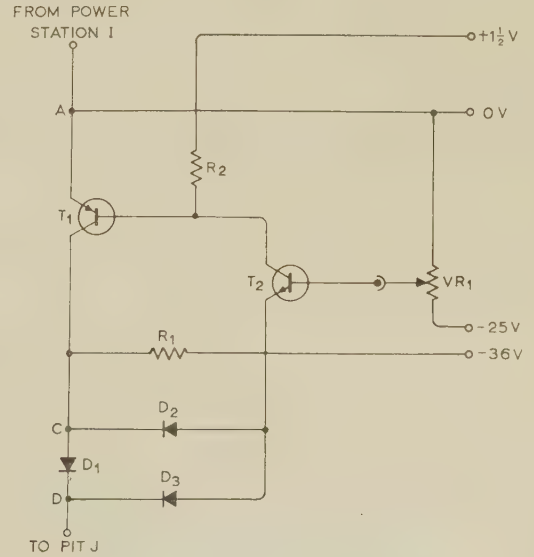


Fig. 5.—Circuit of constant-e.m.f. unit.

T_1, 2N242 or V30/30P R_1, 22kΩ
T_2, 2N35 R_2, 3·3kΩ
VR_1, 50kΩ $D_1D_2D_3$, CV448

on a conventional transistor shunt-stabilized voltage supply. In this case, however, the unit is not required to supply current to the matrix, so the initial current flowing through T_1 needs only to be large enough to establish the required voltage between A and C (V_{AC}). The potentials shown are supplied from dry batteries, and as the 0-volt line must be floating, independent supplies are needed for each row of the matrix. There are 20 e.m.f. units in one row, so that current requirements must be minimized.

The required voltage is set on the potentiometer VR_1. Since the base-emitter p.d. of T_2 is very small under conducting conditions, V_{AC} is virtually equal to the set voltage. The current flowing through R_1 $[=(36 - V_{AC})/R_1]$ is composed of the leakage current through T_1 (which is cut off because R_2 is connected to the $+1\frac{1}{2}$-volt line) plus any extra needed through T_2. Current fed into the circuit from a power station supplies part of the leakage current of T_1, so that T_2 must pass extra current. This continues until the voltage drop across R_2 is such as to bring T_1 into a conducting condition. T_1 is then able to pass any required current up to the design maximum (10 mA) and it, together with T_2, forms a 2-stage amplifier which enables the voltage reference source, VR_1, to be of high impedance.

Thus, with small currents the circuit acts as a single-stage-amplifier voltage-stabilized source so as to minimize initial current demands. As the current passing through the unit increases, the amplifier changes into a 2-stage device to cop with these larger currents. D_2 and D_3 discount the voltag drop across D_1 when conducting. If D_1 is cut off the s voltage is developed between A and C; when D_1 conducts th voltage is established between A and D and the unit then has performance as shown in Fig. 6.

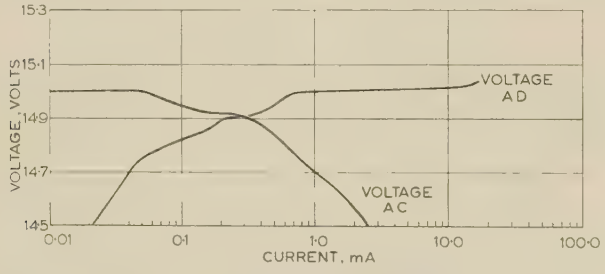


Fig. 6.—Performance of constant-e.m.f. unit.

The matrix itself is only a 5 × 20 set of potentiometers an sockets into which the cost units are plugged. The sockets a interconnected in rows and columns with additional wiring f monitoring purposes.

(6) RESULTS

A typical problem solved on the computer is shown in Table This shows the cost (in pence per ton) of each route for which cost is known, the outputs of the pits and the requirements the power stations.

Table 2 shows the problem as scaled to fit the computer. Th costs have been adjusted as described in Section 4, and t resulting voltages are shown in the top left-hand corner of ea matrix junction. The pits and power stations have their outpu and inputs represented in milliamperes, and paralleling of curre sources to represent power stations with large requirements w be noted. The initial solution of the problem is also shown an is indicated by the figures in parentheses. It can be seen that the are 24 conducting paths, so that with a 5 × 20 matrix the solutio is unambiguous. This solution was obtained by monitorin current flows, and the routes were checked in columns; by doin this the column totals can be checked as readings are taken. Th solution shows discrepancies in the row totals; these are due several causes, including errors in setting the current generator the 'constant' e.m.f. being not absolutely constant and the matr being disturbed when readings are taken. These inaccuracie which are of a type general to all analogue computers, need litt correction and the true solution can be obtained by inspectio

The full correct solution is shown in Table 3. The break-eve costs are shown for the unpriced routes only; they could als have been read off for the other routes if required. It shou

Table 2

PROBLEM AS SET UP ON COMPUTER. THE NUMBERS IN PARENTHESES SHOW SOLUTION

Current source number	Volts	19	8·5		19·5	4·5	25	9	17·5		6	3	13·5	9·5	5·5			8·5	6·5	16·5	
1 9·0 6 9·0 7 8·0 8 8·2 } 34·2			(1·7)			(·3)		(1·7)	(3·1)		(3·2)	(7·3)	(4·4)	(2·1)	(2·7)			(1·8)	(2·3)	(3·0)	
2 3·5		7·5	3		14·5	7·5	3		3 (0·7)	3 (3·1)	12		3	3		8	6	12·5	12·5	3	
3 7·0 9 7·0 } 14·0					3 (5·4)	3			4	3·5 (3·0)	3	3			3	3 (4·6)	3 (1·2)	15	15	4·5	3
4 7·7		5	3	4	7·5		10·5	10·5		10	11	14·5	11·5	5	5	6 (1·3)		15	15		5 (6·5
5 9·4 10 9·2 } 18·6		3 (4·9)		3 (2·1)	18·5		5 (9·6)	3		7·5	7·5	9		4·5		6 (2·0)		3	3		10
	mA →	4·9	1·7	2·1	5·4	0·3	9·6	1·7	3·8	6·1	3·2	7·3	4·4	2·1	2·7	7·9	1·2	1·8	2·3	3·0	6·5
	Current sink number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

Table 3

COMPLETE SOLUTION WITH BREAK-EVEN COSTS

1	342	19	8·5 (17)	3	19·5	4·5 (3)	25	9 (17)	17·5 (37)	7·5	6 (32)	3 (73)	13·5 (44)	9·5 (21)	5·5 (27)	17	17	8·5 (18)	6·5 (23)	16·5 (30)	16
2	35	7·5	3	0	14·5	7·5	3	0	3 (1)	3 (34)	12	0	3	3	0	8	6	12·5	12·5	3	0
3	140	3	0	0	3 (54)	3	5	0	4	3·5 (27)	3	3	0	0	3	3 (47)	3 (12)	15	15	4·5	3
4	77	5	3	4	7·5	0	10·5	10·5	6·5	10	11	14·5	11·5	5	5	6 (12)	6	15	15	5·5	5 (65)
5	186	3 (49)	0	3 (21)	18·5	0	5 (96)	3	6·5	7·5	7·5	9	2·5	4·5	0	6 (20)	6	3	3	5·5	10
100's of tons per week		49	17	21	54	3	96	17	38	61	32	73	44	21	27	79	12	18	23	30	65
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

Pits

noted that, where a break-even cost is marked as zero, it eans that the cost is less than that of the cheapest route in that lumn.

(7) CONCLUSIONS

The computer was built as a prototype to test the soundness the analogy and to develop the circuits. The analogy has en well established by the widely differing problems which ve been solved by the computer. For a larger computer with, v, a 20 × 60 matrix the circuit requirements would be more ingent and the operating facilities would need to be more mprehensive, but there is no obvious reason why either of se requirements should prove impracticable.

(8) ACKNOWLEDGMENT

The work was carried out in the Central Electricity Research aboratories, and the author thanks the Director of the aboratories, Dr. J. S. Forrest, for permission to publish the per.

(9) REFERENCES

(1) HITCHCOCK, F. L.: 'The Distribution of a Product from Several Sources to Numerous Localities', *Journal of Mathematics and Physics*, 1941, **20**, p. 224.

(2) HAYWARD, A. L., *et al.*: 'Minimization of Fuel Costs by the Technique of Linear Programming', *Transactions of the American I.E.E.*, 1958, **77**, Part III, p. 1288.

(3) BERNERS-LEE, C. M.: 'The Use of Computers for Optimal Planning', *British Communications and Electronics*, 1959, **11**, p. 776.

(4) PYNE, I. B.: 'Linear Programming on the Electronic Analogue Computer', *Transactions of the American I.E.E.*, 1956, **75**, Part I, p. 139.

(5) STRINGER, J., and HALEY, K. B.: 'The Application of Linear Programming to a Large-Scale Transportation Problem', International Conference on Operational Research, Oxford, 1957.

(6) CROSSMAN, E. R. F. W.: Unpublished communication.

(7) JEANS, J. H.: 'The Mathematical Theory of Electricity and Magnetism' (Cambridge University Press, 1927). Fifth Edition, p. 321.

621.396.677.3

The Institution of Electrical Engineers
Paper No. 3364 E
Jan. 196(

THE CROSSED-GRATING INTERFEROMETER: A NEW HIGH-RESOLUTION RADIO TELESCOPE

By Prof. W. N. CHRISTIANSEN, D.Sc., N. R. LABRUM, B.Sc., K. R. McALISTER, and D. S. MATHEWSON, M.Sc.

(*The paper was first received* 19*th December*, 1959, *and in revised form* 22*nd August*, 1960.)

SUMMARY

For the first time, a radio telescope has been constructed with sufficient directivity for the production of detailed 'radio pictures' of the brightness distribution on the solar disc at decimetre wavelengths.

The aerial system, which operates at 21 cm wavelength, consists of two long mutually-perpendicular arrays, each made up of 32 paraboloid antennae, 19 ft in diameter, spaced uniformly along a 1 200 ft base-line. The polar diagram of each array is a set of fan-shaped beams, with high resolution in one direction only. A pencil-beam response (with multiple beams) is produced by combining the signals from the two arrays by a phase-switching method. With the dimensions used, these pencil beams are 3′ of arc in diameter and about 1° apart. West–east scanning is provided by the earth's rotation, and the north–south position of the beams can be adjusted by a phase-shifting mechanism. A series of parallel profiles across the sun is obtained; these profiles may be combined to give a 'picture' of the solar disc.

For satisfactory performance, side lobes in the arrays must be kept at a very low level. To achieve this, the current distribution is tapered from the centre to the ends of each array, and very close tolerances are maintained on the phase relationships between the elements. Phase errors due to thermal expansion of the long twin-wire feeder system are avoided by arranging the lines so that all the elements are connected to the receiver through equal lengths of feeder.

The design and construction of the instrument are described in detail. The paper concludes with an account of the techniques used in testing and adjusting the arrays.

(1) INTRODUCTION

It has been known since the early days of solar radio astronomy that the radio emission from the sun includes radiation from localized sources of various types on the solar disc. To study these sources in detail, it is necessary to devise receiving techniques which give sufficient resolving power to separate them from each other and from the general background of thermal radiation.

The instrument described in the paper represents a novel solution to the problem of attaining high angular resolution at reasonable cost. Operating at a wavelength of 21 cm, it provides, for the first time, sufficient discrimination in two dimensions to permit the production of 'radio pictures' of the sun, i.e. detailed maps of the radio-brightness distribution over the solar disc such as Fig. 1(*a*).

At wavelengths near 20 cm, the solar radiation originates mainly in the chromosphere and the inner corona.* Early observations showed that it consists of two main components,[1] one of which remains constant for long periods, and is identified with thermal radiation from the 'quiet' or undisturbed sun. The other component varies slowly from day to day, and its intensity is closely correlated with the occurrence of visible sunspots and of chromospheric plages [bright areas which are visible when the sun is viewed in monochromatic light, corresponding to line emission from hydrogen or calcium atoms; an example is shown in Fig. 1(*b*)]. Measurements during solar eclipses[2, 3] established that the 'slowly-varying' component is in fact, emitted from small regions (with angular dimensions commonly between 2′ and 6′ of arc) located in the lower corona usually over visible sunspots.

The study of these 'radio plages', as they may be called, is of great importance, since it can be expected to lead to a better understanding of the processes taking place in the solar atmosphere, and of the associated optical phenomena. It also has a bearing on the problems of long-distance radio propagation, since the radio plages are sources of the ionizing radiation which maintains the terrestrial ionosphere.

Detailed observations of the sources of the slowly varying component can obviously only be made by means of radio telescopes with very high angular resolving power. The requirements in this respect are quite severe; thus a half-power beam width of 3′ requires an effective aerial aperture of about 1 000 wavelengths, or 700 ft at a wavelength of 21 cm. To attain such high resolution in one direction only, aerials with one very long dimension have been constructed. Covington,[4] at 10·3 cm wavelength, has used a long slotted-waveguide array; and Christiansen and Warburton[5] devised a multi-element or grating interferometer with a 3′ beam width at 21 cm wavelength. This interferometer consists of a row of elements uniformly spaced along a straight line. The array is analogous to an optical diffraction grating; its polar diagram is a family of narrow, but widely-separated, fan-shaped beams. As the earth rotates these drift in succession across the sun, giving strip-wise scans of the solar disc.

The effect of high resolution in all directions can be obtained by combining strip scans of this kind which have been made in a number of different directions. However, this method is cumbersome, and has the further disadvantage that, as a rule, all the necessary data cannot be compiled in a single day. The technique is therefore unsuitable for observing the varying component of the sun's radiation. For this purpose, a true 'pencil-beam' radio telescope appears to be indispensable. Steerable paraboloid aerials of sufficient size are at present out of the question; the receiving system described in the paper, however, achieves the requisite performance by an alternative method. This new instrument may be described as a crossed multi-element interferometer. The aerial consists of a pair of mutually-perpendicular interferometer arrays. High resolution in two dimensions is attained by applying the principle of Ryle's phase-switching interferometer,[7] so that the two arrays are combined to form a type of Mills cross;[8] the system differs from the latter, however, in having a multiplicity of responses instead of a single response. The two arrays are connected to the receiver through a phase-switching circuit, which reverses the relative phases of the two signals at a low frequency. This

* The sun's atmosphere, outside the photosphere, or visible surface, consists of two main regions. The lower layer is the chromosphere, which extends to a height of a few thousand kilometres and has a temperature of the order of 10 000° K. Beyond this is the more tenuous corona, which stretches outwards for several solar radii and consists of highly ionized gas with a kinetic temperature in the region of one or two million degrees absolute.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Prof. Christiansen was formerly at the C.S.I.R.O. Radiophysics Laboratory, and is now Professor of Electrical Engineering at the University of Sydney, New South Wales, Australia.

Mr. Labrum, Mr. McAlister and Mr. Mathewson are at the C.S.I.R.O. Radiophysics Laboratory, Sydney, New South Wales, Australia.

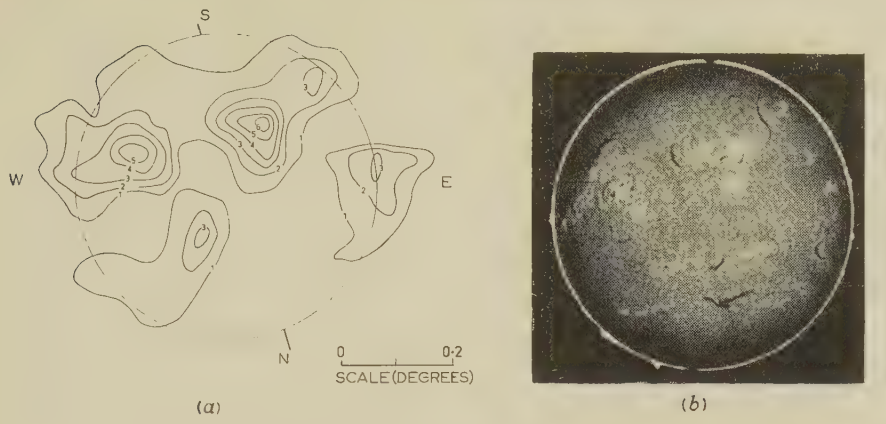


Fig. 1.—Pictures of the sun taken on 2nd December, 1957.

(*a*) Radio isophotes, $\lambda = 21$ cm.
(*b*) Hα line emission, $\lambda = 6{\cdot}56 \times 10^{-5}$ cm.

tching process has no detectable effect on a signal which is ng received in one aerial only. However, when a source is ıated so that signals from it are received in both aerials, the ponses which it produces in the latter are coherent, and the ıse switch causes them alternately to add and cancel. The eiver output thus contains a component modulated at the tching frequency, the amplitude of which is a measure of power received from parts of the sky which are common to h sets of fan beams. If this component is recorded separately m the unmodulated background, the effective polar diagram the whole system is reduced to a set of narrow pencil beams, with maximum response in directions corresponding to the intersections of the fan beams of one array with those of the other.

The dimensions of the aerials are chosen so that these pencil beams are about 1° apart (this spacing is sufficient to ensure that no more than one beam can lie on the sun at any time) and 3′ in diameter to half power. By using the earth's rotation, a succession of pencil beams is allowed to cross the sun along parallel lines, in the manner of a television scan. The distribution of radio emissivity over the solar disc is thus determined in a direct, rapid and unambiguous fashion.

Fig. 2 is a general view of the aerial, which is located at

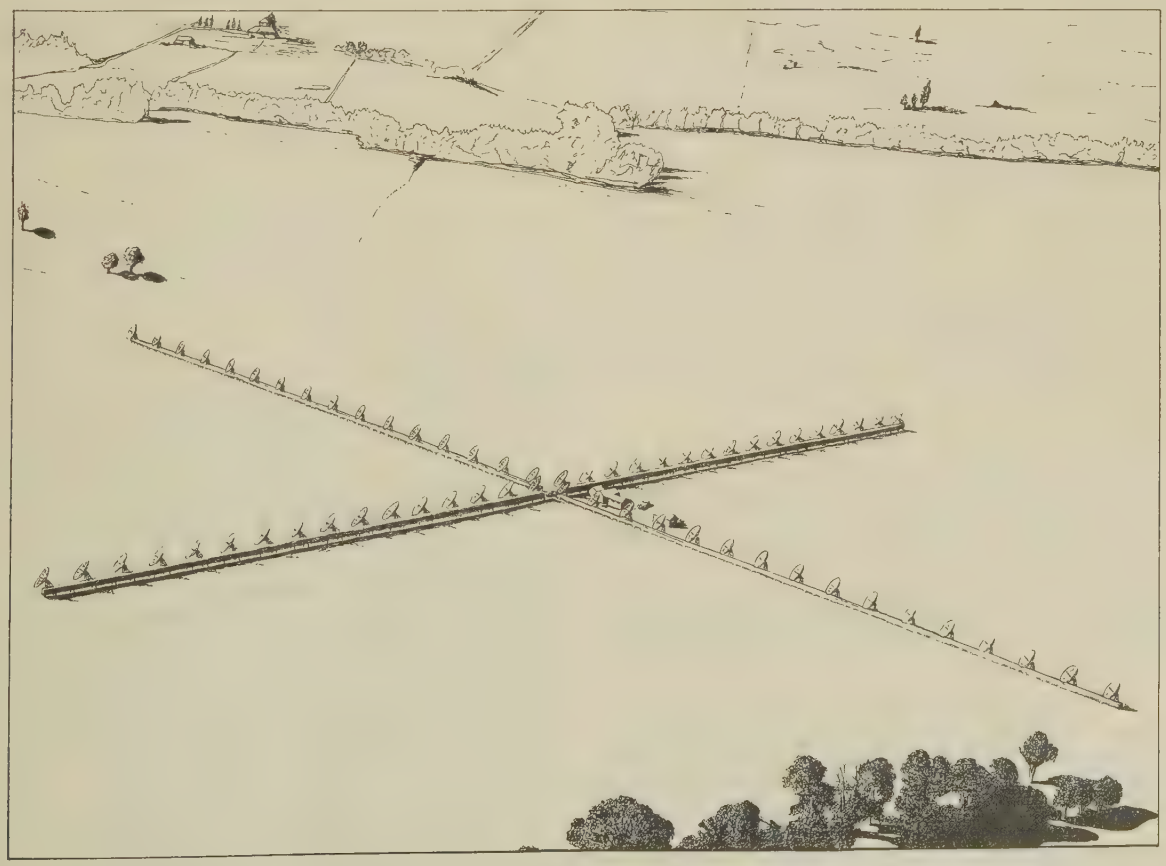


Fig. 2.—The crossed multiple-element interferometer at Fleurs, New South Wales.

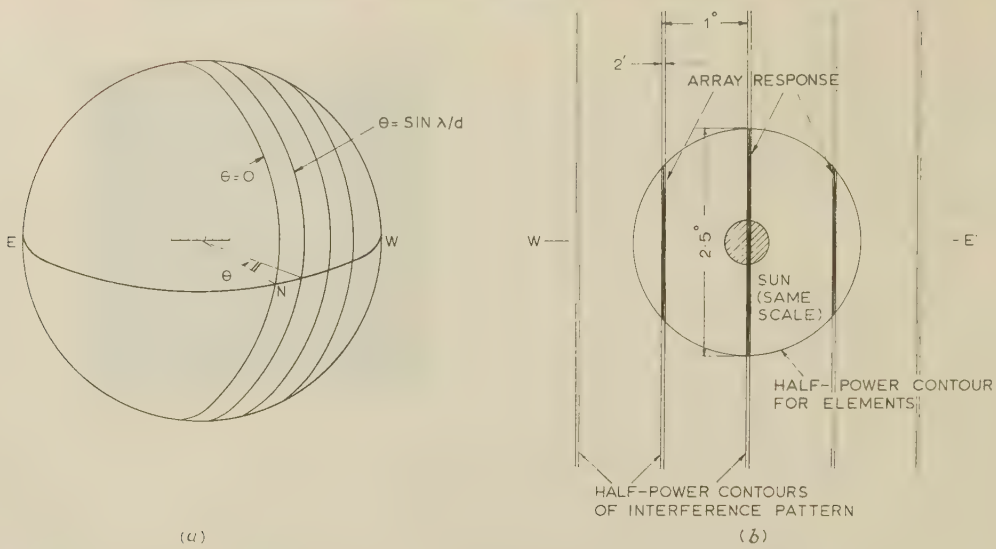


Fig. 3.—Response of a multi-element interferometer.

(*a*) Directions of maximum response for a multi-element interferometer.
(*b*) Response pattern of the east–west interferometer array at Fleurs.

Fleurs, near Sydney, New South Wales. A brief account of the method,[9] and also a preliminary survey of results obtained with it,[10] have already been published.* The object of the paper is to present a discussion of the principles involved in the design of the instrument, together with a detailed description of its construction and operation.

(2) CROSSED MULTI-ELEMENT INTERFEROMETER

(2.1) Aerial Arrays

In its simplest form, the multi-element interferometer consists of a row of identical aerial elements, spaced at equal intervals along a straight line. An array of this kind is analogous to an optical diffraction grating; its directivity (in terms of received power) is given by

$$P(\theta) = P_1(\theta)\,\frac{\sin^2 Ne}{N\sin^2\theta} \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where

θ = Angle between the normal to the system and the direction of the ray.
d = Spacing between adjacent elements.
λ = Wavelength.
$e = \dfrac{\pi d \sin\theta}{\lambda}$
N = Number of elements.
$P_1(\theta)$ = Power received by each element from a point source in direction θ.
$P(\theta)$ = Power received by the array from the same point source.

The second term in this equation represents the interference pattern associated with the array as a whole. Directions of maximum response are given by

$$d\sin\theta = n\lambda \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where n is an integer. On the celestial sphere, the points satisfying eqn. (2) lie on a series of small circles† with their centres along the axis of the array [Fig. 3(*a*)]. The direction from which signals can be received at any time are restricted to the part of this interference pattern contained within an envelope corresponding to the directivity of the elements, as illustrated in Fig. 3(*b*). The elements must therefore be steered so that the source under observation is always located at the centre of this envelope.

(2.2) The Cross

The principle of the phase-switching method adopted for producing a pencil-beam polar diagram has been outlined above. It can be shown[8] that the pencil-beam response to a signal from any given direction is twice the geometric mean of the responses of the two aerial arrays to the same signal

$$R = 2(R_1R_2)^{1/2} \quad . \quad . \quad . \quad . \quad . \quad ($$

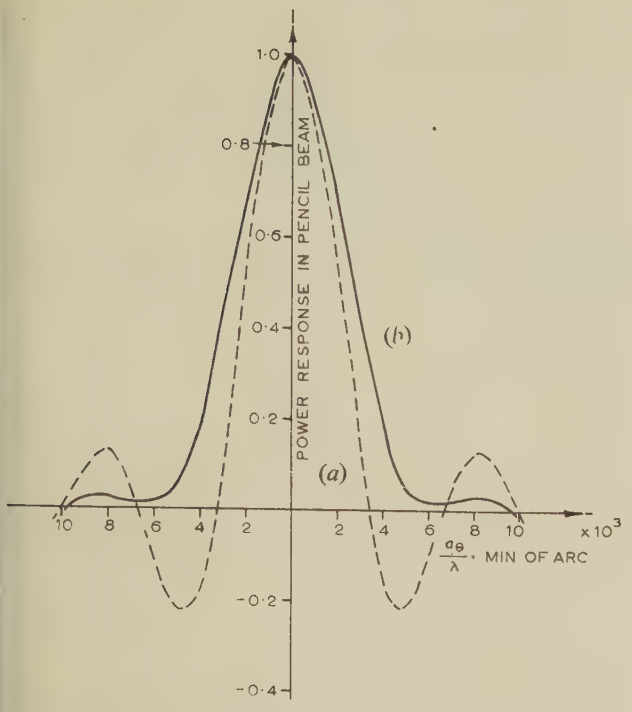
In the present system, the two sets of fan beams are approximately orthogonal over most of the sky. In a plane parallel to, say, one of the beams of the first array, R_1 is constant, so that R is proportional to $\sqrt{R_2}$, i.e. the power polar diagram of the pencil beam in this plane is the same as the voltage polar diagram of the second aerial. This means that the width between half-power points is somewhat greater for the pencil beam than for the fan beams, and also that any side lobes which are present in the responses from the separate arrays are greatly accentuated in the cross presentation. The latter effect is the more serious. For example, Fig. 4(*a*) shows one of the pencil beams when the arrays are uniform multi-element interferometers. The side lobes are up to one-fifth of the main response (in terms of received power). This is unacceptable, and it is necessary to reduce the side lobes. In discussing this, it is simplest to consider the equivalent case of the same aerial when used for transmission. The desired effect can be obtained by tapering the currents in the elements down from the centre to the ends of each array. The result, with the particular current distribution adopted for this instrument, is shown in Fig. 4(*b*). Tapering inevitably produces some broadening of the main lobe.

(2.3) Scanning Technique

The crossed multi-element interferometer has a polar diagram consisting of a pattern of narrow pencil beams with approximately uniform spacing. As the earth rotates, a succession

* The daily maps are being published in the International Astronomical Union's *Quarterly Bulletin on Solar Activity* and in the *Annals of the International Geophysical Year*.

† A small circle on a sphere is one the centre of which does not coincide with the centre of the sphere. A familiar example is a terrestrial line of latitude (other than the equator).



ig. 4.—Polar diagrams for the pencil-beam response at the zenith of a multi-element interferometer.

(*a*) ---- Uniform arrays.
(*b*) —— Arrays with current distribution

$$I_n \propto \left(0{\cdot}2 + 0{\cdot}8 \cos^2 \frac{\pi x_n}{a}\right).$$

a = Aperture of either array.
θ = Angular displacement from a direction of maximum response.

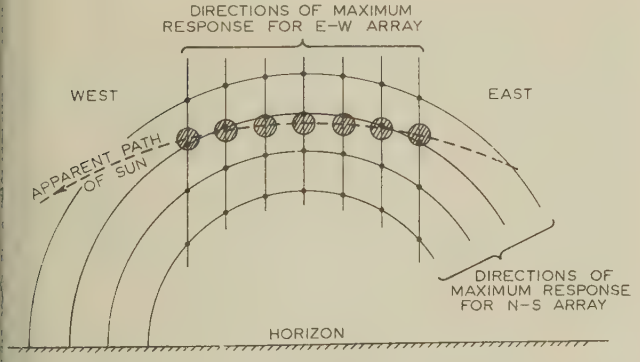


Fig. 5.—Movement of the sun across the beams of the crossed interferometer.

he shaded circles represent successive positions of the sun, and the dots correspond he pencil beams. The vertical displacement between successive beams is greatly ggerated in this diagram.

se passes across the sun. This scanning action is illustrated Fig. 5, where, for convenience, the aerial is assumed to be tionary whilst the sun passes through the pencil-beam systems m east to west.

The geometry of the system is such that successive beams ss the solar disc at progressively changing declinations. It therefore possible to obtain the data needed for a complete -dimensional 'picture' of the sun simply by recording the eiver output whilst a succession of pencil beams drift across solar disc. The scanning process can, however, be carried t more quickly and conveniently if the beams are shifted in clination, so as to maintain a space about equal to the beam width between adjacent scans. This adjustment is made by means of a phase-shifting mechanism in the north–south interferometer.

(3) CHOICE OF EQUIPMENT PARAMETERS

(3.1) Wavelength

One of the main objects of this work was to study the slowly varying emission from the radio plages; this is most prominent at wavelengths between 10 and 30 cm. The wavelength adopted was 21·1 cm ($f = 1{\cdot}423$ Gc/s), so that, if required, the aerial could later be used in investigations of hydrogen-line emission at 1·420 6 Gc/s.

(3.2) Aerial Aperture and Taper

To resolve details on the solar disc, a pencil-beam width of about 3′ is necessary; response in any side lobe should not exceed 3% of that in the main beam. Analysis showed that the side-lobe requirement could be satisfied with a current distribution (if the aerial were used for transmission) given by

$$I_n \propto 0{\cdot}2 + 0{\cdot}8 \cos^2 \frac{\pi x_n}{a} \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where a = Total length of the array.
x_n = Distance from the centre to the nth element.
I_n = Current in that element.

Fig. 4(*b*) shows the polar diagram of the pencil-beam responses when the two arrays are tapered in this way. The maximum side-lobe level is 2½%, and the width between half-power points is (at the zenith) $5\,400\lambda/a$ minutes of arc.

For a beam width of 3′, an aperture of $1\,800\lambda$ or 380 m (about 1 200 ft) is required. In directions away from the zenith, the beam width is increased by the foreshortening of the arrays. This deterioration can be almost eliminated in the case of the east–west array by restricting the observations each day to a period within about two hours of transit; some broadening of the north–south array beams must, however, be accepted in the winter months. In the latitude of Sydney, the zenith distance of the sun at transit is 56° at the winter solstice. The effective aperture of the north–south array in that direction is $1\,030\lambda$, and the width of the pencil beam in the north-south plane is then 5·4′. If the receiver bandwidth is excessive, a further loss of directivity occurs at large zenith distances, because the directions of maximum response are dependent on frequency. It follows from eqn. (2) that, for a small frequency change δf, the shift $\delta\theta$ in the position of a particular beam is

$$\delta\theta = \tan\theta \frac{\delta f}{f} \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

The value chosen for the bandwidth is a compromise, since reducing it also reduces the receiver sensitivity.* A value of 300 kc/s was adopted; from eqn. (5), this spreads the aerial beam by a further 1·1′ at zenith distance 56°, so that, for mid-winter solar observations, the pencil beam is 6·5′ wide in the north–south plane. At midsummer the corresponding beam width is 3·1′.

(3.3) Spacing between Aerial Elements

In order to avoid ambiguity due to reception of solar radiation simultaneously in more than one of the pencil beams, the separation, θ_1, between the directions of maximum response must be made greater than the angular diameter of the sun.

* When, as in this case, the input signals have the characteristics of random noise, the ratio of the output signal to the statistical fluctuations in the receiver output increases with increasing bandwidth (e.g. see Reference 11).

The latter is about 0·6° at 21 cm wavelength, and the minimum value of θ_1 was accordingly arranged to be 1°. θ_1 is determined by the spacing between aerial elements:

$$\theta_1 = \frac{\lambda}{d \cos \theta} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

The required spacing, d, is 58λ (12·3 m or 40 ft). With the aperture already fixed at $1\,800\lambda$, this requires arrays of 32 elements each.

(3.4) Aerial Elements

The dimensions of the individual aerial elements are determined primarily by sensitivity considerations. As a fairly arbitrary criterion, it was decided that the overall sensitivity of the system should be such that the pencil-beam signal due to quiet sun radiation alone would give a receiver output equal to 15 times the r.m.s. noise fluctuations.

If the brightness temperature of the source is T_b, the received power per unit bandwidth is (for a lossless plane-polarized aerial) given by

$$P_1 = \frac{k}{\lambda^2} \int T_b A d\Omega \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

where A = Aerial receiving area, in the direction of the element $d\Omega$ of solid angle.
k = Boltzmann's constant.

If one pencil beam of the crossed interferometer response pattern is directed at the centre of the solar disc, T_b is constant over this beam and zero over all others. The last equation can then be rewritten as

$$P_1 = \frac{kT_b}{\lambda^2} \int A d\Omega \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

the integration now being made over the central pencil beam only.

The integral in eqn. (8) can be evaluated graphically from Fig. 3(*b*), in terms of A_0, the receiving area in the direction of maximum response. For uniform arrays, A_0 would be equal to the sum of the receiving areas of all the elements; the tapering used reduces it by a factor of 0·8 : 1. The minimum value of A_0, and therefore of the receiving area of each element, can be determined by making P_1 equal to 15 times the received power required to produce output equal to noise fluctuations. Christiansen and Warburton[6] showed that the quiet sun temperature at the centre of the disc is (at sunspot minimum) 5×10^4° K. Attenuation in the aerial feeders was taken as 10 dB; for the receiver, the noise factor was assumed to be 10 dB, the bandwidth was 300 kc/s, and the output time-constant 0·5 sec. With these values, the receiving area for each element is found to be 21 m^2.

At 21 cm wavelength, the most convenient form of aerial is a dipole-fed paraboloid reflector. If the efficiency of such an aerial is taken as 0·5 (a fairly low estimate), the required reflector diameter for the elements of the interferometer arrays is 5·8 m (19 ft).

(4) DESIGN

(4.1) Aerials

In designing the aerials and feeders, which together account for most of the cost of the instrument, it was important to aim at the cheapest possible construction consistent with the required performance and durability.

One of the 64 aerials is shown in Fig. 6. The tower is made from steel tubing; its four feet are bolted to concrete blocks set in the ground. It carries an equatorial mounting* for the paraboloid reflector and dipole feed. The reflector consists of ½ in-mesh galvanized wire supported by an aluminium tube framework. Its diameter and focal length are 19 ft and 6 ft 3 in respectively. The dipole and its reflecting plate are at the end of a metal tube which is mounted at the pole of the mirror. Connection to the dipole is made through a twin-wire line inside the supporting tube.

The aerials are fitted with scales of declination and hour angle, to facilitate adjustment. The declination is set manually (see Fig. 6) by means of two nuts which run on a semicircular screw and turn the aerial assembly about an axis at right angles to the polar axis.

In order to follow the sun across the sky during a day's observing period, it is necessary to drive all the aerials in hour angle, i.e. in rotation about the polar axis. This might be done by using a separate driving mechanism for each aerial. However, it was found to be much more economical to move all the aerials by means of a central control system. A hydraulic ram pulls a ⅜ in-diameter steel cable, which extends the whole length of an array and is kept taut by large weights at the ends. At each aerial this cable is clamped to a pivoted drive arm which advances a ratchet wheel by one tooth for each to-and-fro movement of the cable. This drives the aerial about its polar axis. Each stroke changes the hour angle by ⅛°, which corresponds to the earth's rotation in 30 sec of time.

To avoid introducing large switching transients in the a.c. mains supply, a hydraulic accumulator is used. The ram which moves the aerial drive cable is actuated through an electrically controlled valve, which, when opened, connects the ram to the main hydraulic reservoir.

The operation of this valve is usually arranged so that the cable makes one stroke every 30 sec. This ensures that, during the passage of the sun across one of the interference beams, the latter is always very close to the direction in which the elements are pointed. A slightly superior method is to drive the aerials midway between scans, in predetermined steps of about 1°, so that throughout each scanning period the elements are directed exactly at the hour angle of the pencil beam which is crossing the sun. The extra error introduced by the use of the uniform drive is so small, however, that the adoption of the more complicated alternative has not been justified.

The aerials are set up with the prescribed spacing along accurately surveyed lines lying north–south and east–west respectively, intersecting at the mid-points of the arrays. The positions are accurate to within $\pm\frac{1}{8}$ in, both horizontally and vertically.

(4.2) Feeders

The design of the feeder system is based on that used in the earlier multi-element interferometer.[5] Balanced open-wire transmission lines are used; besides being cheap to construct, they are easily accessible for testing, and lend themselves to the type of connection needed in this equipment. At decimetre wavelengths, moreover, their attenuation is only slightly more than that of much more bulky and expensive feeders such as large-diameter coaxial lines. The conductors are hard-drawn copper wires, 0·16 in in diameter, spaced 0·75 in apart between centres. Each wire is held under tension by means of springs. The lines are supported by polythene insulators at 18 ft intervals, and

* In this type of mounting, the aerial is pivoted on a polar axis, i.e. one parallel to the earth's axis. By rotating the aerial about this axis at the correct rate, the direction in which it is pointed can be made to follow closely the apparent diurnal motion of the sun across the sky.

† These are the geocentric (equatorial) co-ordinates ordinarily used to specify the apparent positions of astronomical bodies. Declination, which is analogous to geographical latitude, is angular distance from the celestial equator. Hour angle, corresponding to longitude, is the angular distance westwards from the observer's meridian.

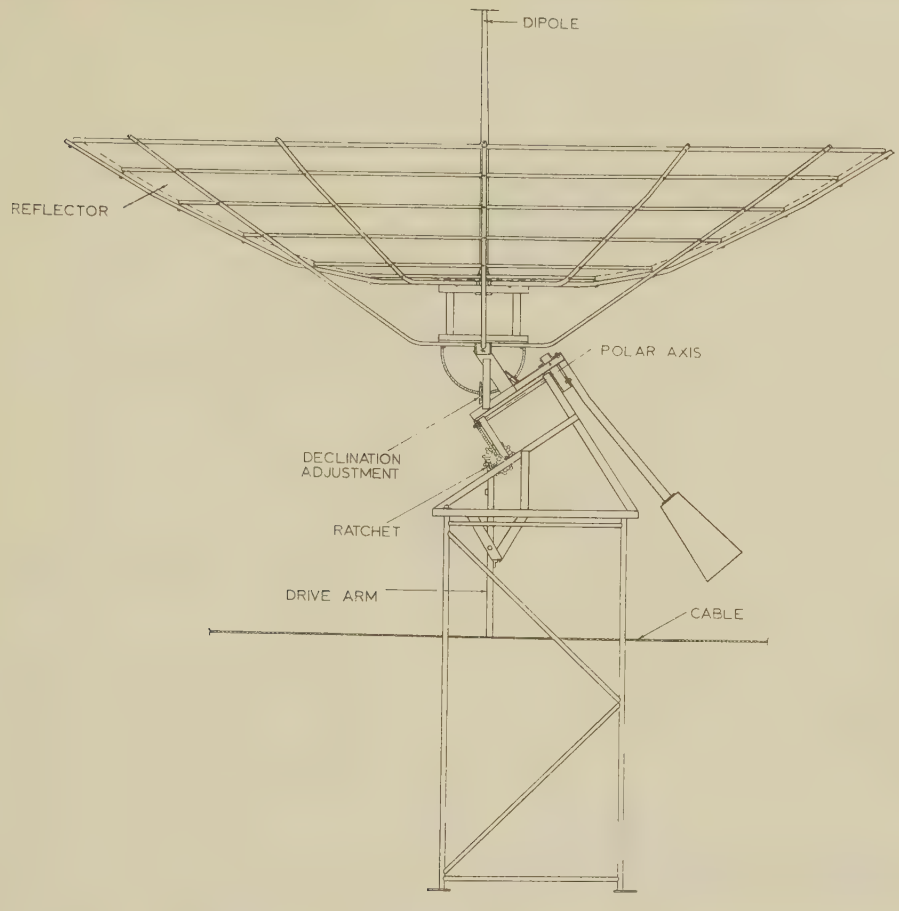


Fig. 6.—An aerial unit, showing the drive mechanism.

tance between conductors is maintained by small polythene acers about every 6 ft. Both the spacers and supports are ounted in pairs with quarter-wavelength spacing, to minimize lections. The characteristic impedance of this line is 270 ohms; e calculated attenuation (for pure copper wires) is 0·8 dB per 0 ft. Measurements made after the lines had been exposed the weather for some months indicated that the loss was then, fact, about 1 dB per 100 ft.

Because of the great electrical length of the feeder lines (the termost aerials in each array are over 900 wavelengths from e receiver), care must be taken to prevent phase shifts between e signals from the various elements due either to thermal pansion of the lines or to changes in atmospheric humidity. These effects are minimized by using a branching system of transmission lines with equal lengths of line between all the aerials and the receiver. Any change in the electrical length thus affects all the aerials equally. The arrangement is shown in Fig. 7. Five twin lines, mounted one above the other for the length of the array, are used. The lines are 6 in apart in the vertical plane. With this spacing, no undesirable effects due to interaction between them can be detected.

The aerials are connected in pairs to sections of the top line. Each of these sections is joined at its mid-point to the line below, etc. The unused portions of feeder are blocked off by short-circuiting plates, as shown in Fig. 8. Each 'active' section of line is supported by springs, located so that changes in

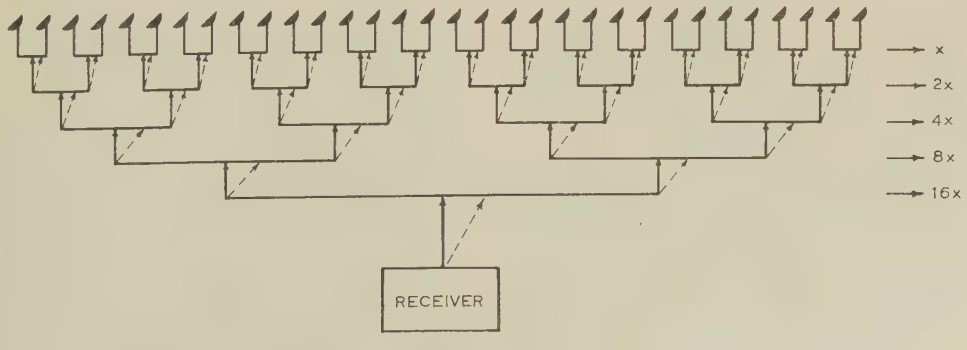


Fig. 7.—The branched feeder system.

The arrows indicate the movements needed in all the sliding junctions at each level in order to change the relative feeder lengths for adjacent aerials by an amount $2x$.

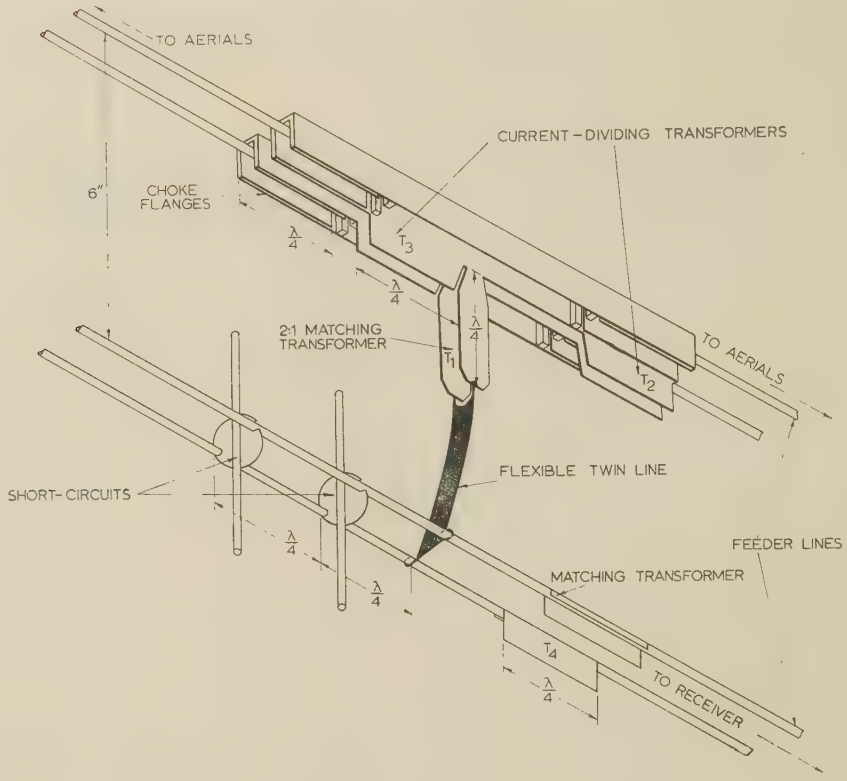


Fig. 8.—A sliding junction in the twin-wire feeder system.

length take place symmetrically with respect to the junction, which therefore has no tendency to shift as the temperature varies.

The connection to each aerial is made through a vertical twin line. Behind the paraboloid, a section of plastic-insulated flexible twin feeder is used, to allow for movement. This is joined to the line which passes up the support tube to the dipole.

The overall length of feeder from the receiver to each aerial is about 600 ft; this includes five junctions between lines. The attenuation between the receiver and a point near the connection to an aerial was found by measurement to average 9·5 dB. Resistive losses in the lines themselves accounted for 6 dB of this; the rest is due mainly to radiation near the junctions (a rough calculation indicates that the radiation resistance for each of the two discontinuities at a junction is about 10 ohms).

(4.3) Junctions

One of the T-junctions used in the branching feeder system is shown in Fig. 8. In addition to making the connections between the lines, these have two other functions. They determine the current division between the various branches, to give the prescribed distribution along the array; and, in the north–south array, they are movable so that the position of the interference pattern can be adjusted by phase-shifting.

The junction fits on the upper line. An arrangement of choke flanges is incorporated, so that its electrical properties are not dependent upon the quality of the mechanical contact between it and the conductors. The lower line is connected to the junction by a 300-ohm flexible twin lead, and a quarter-wavelength transformer, T_1. This consists of a pair of parallel brass plates, and corrects the 2 : 1 mismatch at the junction.

The required current division is produced by the two equal quarter-wave transformers, T_2 and T_3, formed by projecting flaps on the arms of the junction. These transform the impe dances of the matched upper lines to higher and lower value respectively, at the mid-point where they are connected i parallel. In this way current ratios of up to 2·4 : 1 betwee the arms are produced with only slight variations in the impe dance presented at the junction. Small residual mismatches an corrected by an additional matching section, T_4, on the lowe twin line.

(4.4) Phase-Shifting System

For convenient operation of the instrument, the position of th north–south interference pattern must be made adjustable. progressive variable phase shift may be introduced between th aerials of the array for this purpose, by moving all the feed junctions simultaneously in an appropriate manner. Thus, refe ing to Fig. 7, if the movements are, from top to bottom lin x, $2x$, $4x$, $8x$ and $16x$, respectively, towards the north, the lengt of feeder between a given aerial and the receiver is increase by $2x$ in relation to that for the next aerial to the north. change of one wavelength between adjacent aerials shifts th interference pattern by an amount equal to the interval betwee successive maxima, i.e. about 1°.

The phase-changing mechanism is shown in Fig. 9. A cab extends along the array; it is held taut by weights at the end and can be pulled to and fro by a hydraulic ram which is con trolled by a solenoid valve. This cable carries a series of block which shift the junctions by actuating a ratchet mechanism. is arranged that one stroke of the cable moves each junction distance corresponding to a change of $\lambda/32$ in the relative length of the feeders for adjacent aerials. This shifts the interferenc pattern through an angle somewhat less than the width of th pencil beam. When a junction has moved half the wavelengt from its extreme southerly position, a pin on the ratchet whee momentarily disengages the pawl, and allows a spring to retur

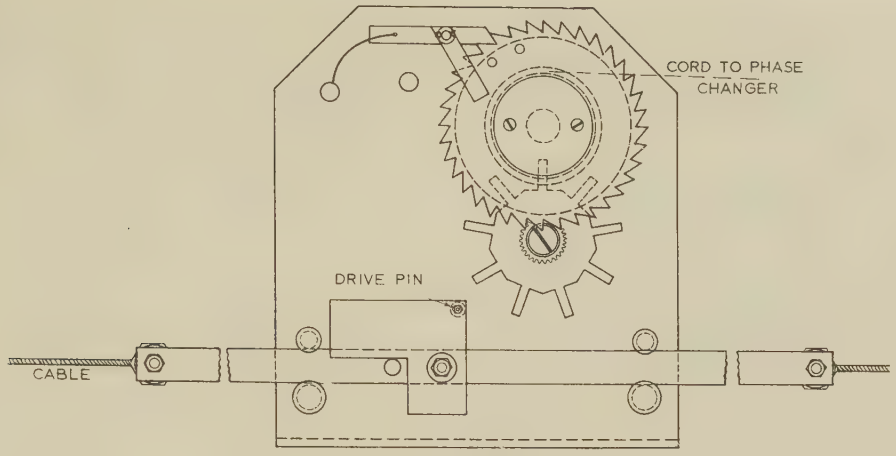


Fig. 9.—Phase-changing mechanism.

junction to its starting point (changes in junction positions integral multiples of $\lambda/2$ do not affect the aerial pattern).

(4.5) Receiver

The signal from each of the aerial arrays first passes through a ance-to-unbalance transformer. The output from the east–st array is taken through a coaxial-line phase switch, which itches the signal between two paths differing by half a wave-gth, at 25 c/s. The east–west and north–south signals are ally added together at a point where they are in either phase anti-phase, according to the switch position.

Following this switching circuit is a transmission-type resonant ity, tuned to 1·423 Gc/s, for image-frequency rejection. The erheterodyne receiver has at its input a crystal mixer, which a 1·393 Gc/s heterodyne oscillator is also coupled. e intermediate-frequency amplifier is tuned to 30 Mc/s, with 0 kc/s bandwidth. The second detector has a square-law aracteristic (this is necessary in order to make the 25 c/s tput independent of the total 30 Mc/s input to the amplifier, ich varies with the power received by the fan-beam systems a whole).

The output is applied to a low-frequency amplifier and then a phase-sensitive detector. In this way, the envelope of the c/s modulation produced by the phase-switching process is lated and rectified. The final d.c. output is therefore propor-nal to the amplitude of this modulation, which, as has been plained, is a measure of the power in the pencil beams formed the intersecting interference patterns. This signal is applied a pen recorder.

(5) ADJUSTMENT AND TESTING

(5.1) Tolerances

In setting up the interferometer arrays, great care is needed ensure that the elements are accurately in phase, and that the riation in amplitude along the arrays is of the required form. ie most serious effect of maladjustment, particularly with the oss-presentation, is the production of side lobes. If one ment has an error, either in phase or amplitude, the overall lar diagram of the array is the pattern shown in Fig. 4(*b*), gether with an 'error polar diagram' due to the error response the element in question.

In the present aerial system phase errors are the most serious use of side lobes. Effects due to aerials near the centres of e arrays are particularly troublesome. This is due partly to e taper in current amplitude, and partly to the fact that the error polar diagram is very broad in the case of a central aerial, so that such side lobes can accept radiation simultaneously from a large part of the sun. In practice, it was found that the aerial pattern deteriorated appreciably if errors of more than 6° occurred in the phases of any of the elements.

It was found necessary to keep errors in aerial-current amplitudes below $\pm 10\%$. The current in a given aerial is determined by the current divisions at five junctions in the feeder system (see Fig. 7). Each of these, in turn, depends on the setting of the transformers T_2, T_3 (Fig. 8) and also on the matching of the two upper arms of the junction. The required accuracy can be attained if the lines are all matched to voltage standing-wave ratios not greater than 1·03 : 1, and the actual current ratio at each junction is then adjusted to within $\pm 2\%$ of the nominal value.

(5.2) Matching

For testing purposes, it was found convenient to treat the arrays as transmitting aerials, and to feed power into them from a test oscillator through what are normally the receiver connections. For impedance measurements, a standing-wave indicator of the type shown in Fig. 10 was used. Two brass probes were

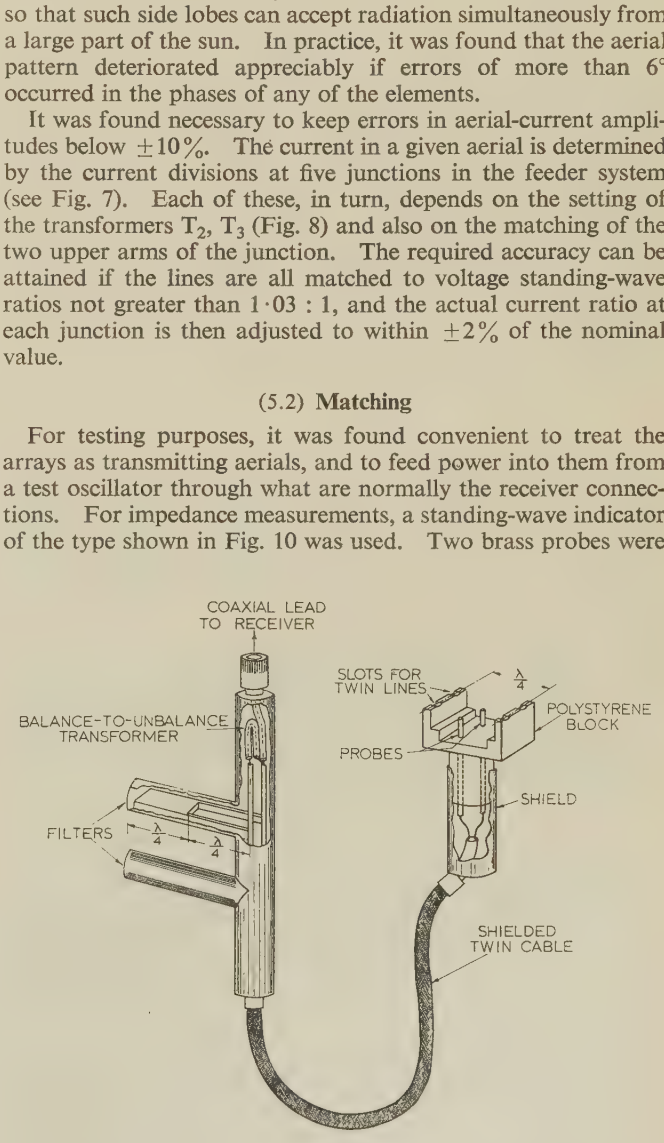


Fig. 10.—Cut-away diagram of standing-wave detector.
The filter stubs remove any residual unbalance from the balanced line.

embedded in a block of polystyrene, the end of which was machined into two bars which fitted on the twin line $\lambda/4$ apart. The probe was connected, through flexible lines and a balance-to-unbalance transformer, to a mobile radio receiver. The receiver output was used as an indicator of the r.f. voltage on the line to which the probe was coupled.

The individual aerials were first matched to the transmission lines. Large reflections were removed by adding $\lambda/4$ transformers, similar to T_4 in Fig. 8, and consisting of pairs of copper plates soldered to the lines. The correct position for the matching section was in each case found by a standing-wave measurement; the plates were then hung on the line and their spacing varied until a match was attained. Small mismatches (v.s.w.r. $< 1{\cdot}05$) could be corrected by slight adjustment of the positions of the spacing beads.

After each pair of aerials had been matched, the junction connecting them to the second of the twin lines was adjusted for current ratio, as explained in Section 5.3. The matching process was then continued through the branching feeder system until the central junction was reached.

(5.3) Current Division

The current dividing ratio at each junction was set after matching the feeders leading from it towards the aerials. The probe and receiver were used to measure the currents in the upper lines; for this purpose, the second-detector current meter was calibrated in terms of receiver input power by means of a signal generator fitted with a calibrated attenuator. When necessary, the transformers T_2 and T_3 were adjusted by bending the $\lambda/4$ plates. The overall effect of these adjustments was then checked by measuring the relative current in each of the aerial feed dipoles. To do this, a pick-up dipole connected to the receiver input was held at a fixed distance from each of the aerial feeds in turn and the receiver responses were observed.

(5.4) Phase

The position for each junction was initially determined by short-circuiting the aerials and then finding the electrical centre of the section of line by observing voltage minima. Final adjustments to bring all the aerial currents accurately into the same phase were made after the matching and current-division measurements had been completed. For this purpose, the phase-comparison system shown in Fig. 11 was used. The output from the test oscillator was divided at a T-junction. Part of the power was fed into the array, as before; the remainder passed through a long flexible coaxial cable and a section of telescopic line to the mobile receiver. A second approximately equal signal was fed into the receiver through another flexible cable, from the pick-up dipole.

With the test dipole held in its standard position on the end aerial of an array, the telescopic line was adjusted until the receiver output was at a minimum, i.e. the signals arriving by the two paths were in anti-phase. The process was repeated for the next aerial, and, if necessary, the appropriate junction was moved along until the two aerial signals were in exactly the same phase. Four adjoining aerials could be compared directly in this way, and the comparison could easily be extended in steps to cover the whole array.

It was found that, with proper care, this method gave results which were reproducible to within $\frac{1}{8}$ in. For maximum accuracy, it was important to avoid twists or large strains in the flexible cables. It was also necessary to carry out the phase measurements at times when the ambient temperature was fairly constant, in order to avoid errors due to differential expansion in the lines.

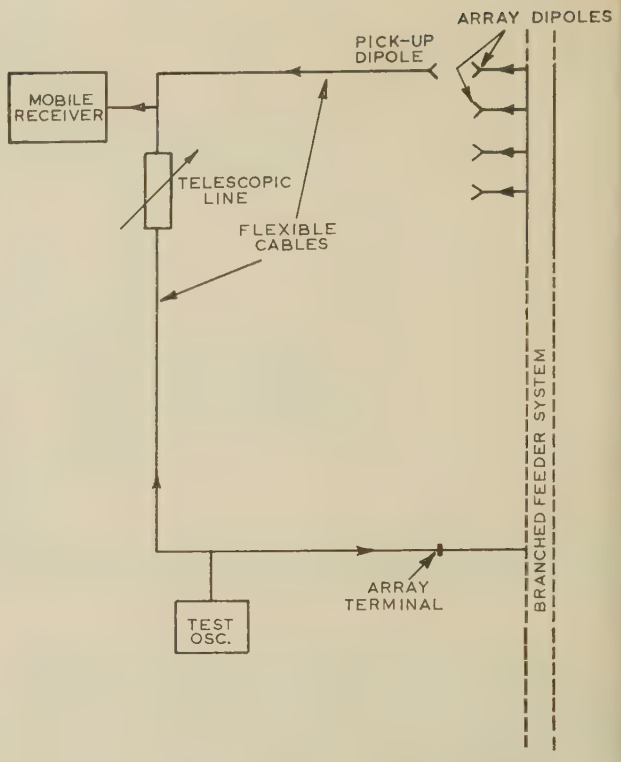


Fig. 11.—Phase comparator, for aerial adjustments.

(6) OPERATION

The crossed interferometer has been in continuous use sinc July, 1957. The expected performance has been fully mai tained throughout this period, with side-lobe levels and outp signal/noise ratio at approximately the predicted values. It h been found that aerial matching and phasing need readjustme only at intervals of about six months. The aerial respon appears to be completely unaffected by thermal expansion i the feeder system. The instrument cannot, however, be used i heavy rain, since the transmission lines are partly exposed to th weather. This is not a very serious limitation, as the aerial located in an area which has a fairly dry climate, with no snowfa and little frost. Interruptions due to rain have been minimize by building roofs over the feeders.

The interferometer has been used mainly in the preparatio of daily maps of the radio brightness distribution over the sola disc. Observations for this purpose are usually begun short before the sun's time of transit, and occupy about an hour.

During the observing period, the aerial elements are steere so that their beams are always pointed approximately at th centre of the sun's disc. The phase-changers in the north–sout array are initially adjusted so that at transit one of the pen beams will pass just south of (i.e. above) the sun (see Fig. 5 The north–south phasing is progressively changed in a pr determined sequence, so that adjacent scans are kept about apart in declination. Without adjustments, the scans wou be too far apart, except during a short period near transit.

A typical record of the receiver output is shown in Fig. 12(a It consists of a series of radio-brightness profiles on parallel lin across the solar disc.* To produce a complete map of the su these profiles must be fitted together in their correct relati positions [see Fig. 12(*b*)]. The calculations, though straigh

* The signal phase changes by 180° between successive lobes of the interfere pattern; consequently the deflections in the pencil beam record are reversed in alterna scans. For the sake of clarity, the record in Fig. 12(*a*) has been redrawn witho these reversals.

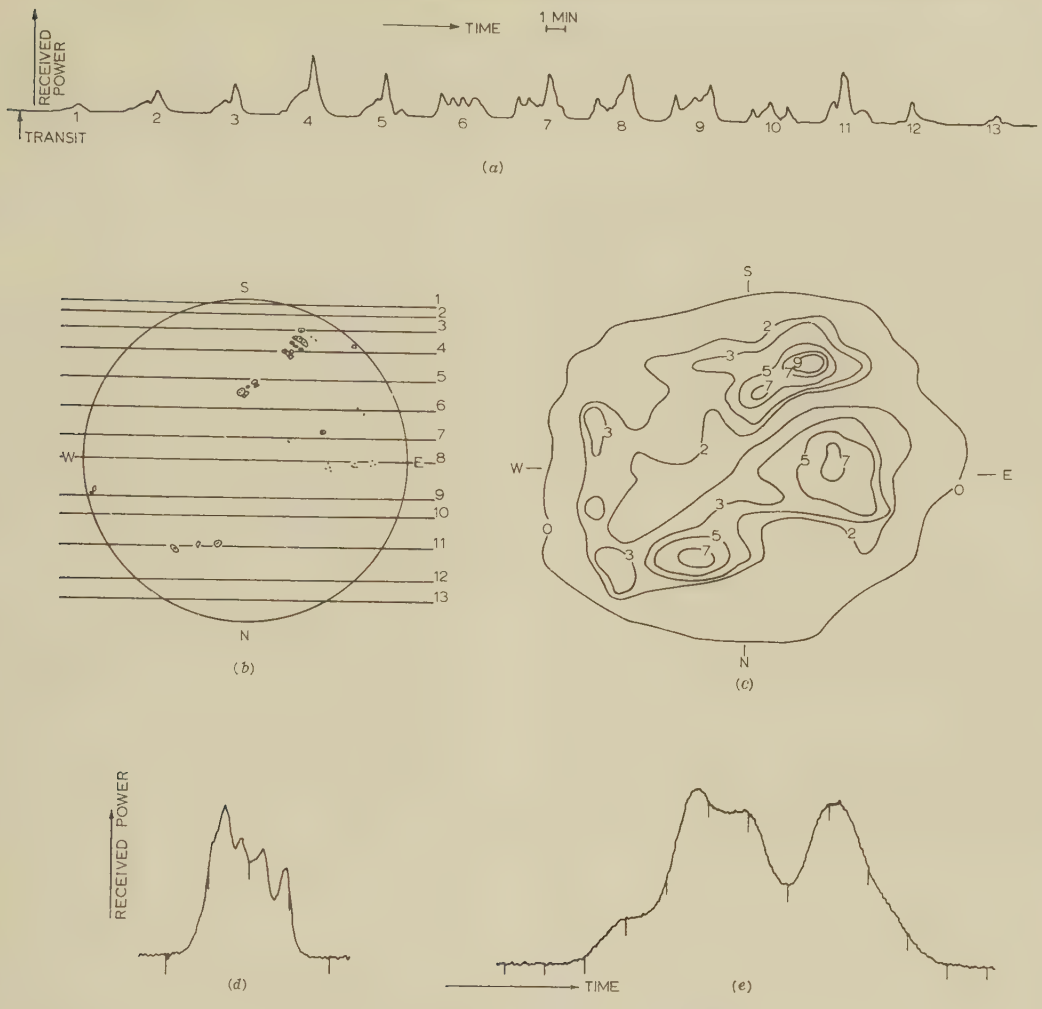


Fig. 12.—Typical radio picture of the sun (22nd October, 1957).

(*a*) Record of received power, as a series of pencil beams crossed the sun.
(*b*) Location of the profiles on the sun (numbered to correspond with the record). Positions of visible sunspots are also shown.
(*c*) Brightness contours derived from (*a*) and (*b*). The unit of contour height is 100 000 °K in brightness temperature.
(*d*) One-dimensional scan with east–west array.
(*e*) One-dimensional scan with north–south array. On the last two diagrams, 1 min time marks are shown.)

iward, are laborious; to shorten them, a table has been pared, giving the co-ordinates of the centres of all the pencil ams. Since the position of the sun at any instant is known, beam which corresponds to a particular scan can readily be ntified, and its exact position on the sun determined from the le. The next step is to convert the vertical deflections in g. 12(*a*) into absolute values of received flux density. A mplete daily calibration of the receiving system would be practicable, as it would involve a determination of all losses, luding those in the long feeder system. An indirect method therefore used. The flux density from the whole sun (in itrary units) is found by a numerical integration of values d from the profiles. The result is compared with a simul- eous absolute measurement of the total flux density, made with eparate simple radiometer at the same wavelength. A scaling tor is thus obtained, so that deflections on the interferometer ord can be converted to absolute flux densities. These are n plotted (as equivalent source brightness temperatures) in form of a contour diagram [Fig. 12(*c*)].

Provision is also made for separate recording of the output m a single array. Either of these signals represents a one-dimensional strip scan across the sun; typical examples are shown in Figs. 12(*d*) and (*e*). The output from the east–west aerial consists of a sequence of scans from west to east across the disc, repeated at approximately 4 min intervals as successive fan beams cross the sun. It is thus particularly suitable for the observation of rapid variations in radio emission of the sun, such as are often associated with solar flares.

(7) CONCLUSION

The instrument has demonstrated that it is feasible to construct decimetre-wavelength receiving aerials with apertures of the order of 2000 wavelengths, and to combine arrays of these large dimensions for the production of pencil beams with widths as small as 3′ of arc.

It has been found possible, without any sacrifice of performance, to use an extremely economical design. Major savings in cost were achieved by using a single mechanism to steer all the 64 aerial elements, and by adopting the simple but efficient twin-wire feeder system.

This is the first aerial of its kind to be built; it is likely, however, that similar and even larger aerial systems will eventually be

fairly extensively used. It is hoped that the experience gained with the equipment described in the paper may be of assistance in future designs.

(8) ACKNOWLEDGMENTS

The authors wish to thank Mr. K. W. Anderson, who was responsible for a considerable part of the mechanical design of the aerial system, and also the team, led by Mr. E. C. Chenhall, who very ably carried out the construction and installation. Since the completion of the instrument, Messrs. E. Harting, M. W. Willing and C. Turrall have given much assistance in both maintenance and observations. Thanks are also due to Mr. D. E. Yabsley for assistance in the preparation of the paper. Finally the authors wish to acknowledge gratefully the many valuable discussions they have had with Dr. J. L. Pawsey during the course of the project, and the constant support and help which he has given to the work.

(9) REFERENCES

(1) Pawsey, J. L., and Yabsley, D. E.: 'Solar Radio-Frequency Radiation of Thermal Origin', *Australian Journal of Scientific Research*, A, 1949, **2**, p. 198.

(2) Christiansen, W. N., Yabsley, D. E., and Mills, B. Y.: 'Measurements of Solar Radiation at a Wavelength of 50 cm during the Eclipse of November 1, 1958', *ibid.*, 1949, **2**, p. 506.

(3) Covington, A. E., Medd, W. J., Harvey, G. A., and Broten, N. W.: 'Radio Brightness Distribution of the Sun at a Wavelength of 10·7 cm, June 30, 1954', *Journal of the Royal Astronomical Society of Canada*, 1955, **49** p. 235.

(4) Covington, A. E., and Broten, N. W.: 'Brightness of the Solar Disk at a Wavelength of 10·3 cm', *Astrophysical Journal*, 1954, **119**, p. 569.

(5) Christiansen, W. N., and Warburton, J. A.: 'The Distribution of Radio Brightness over the Solar Disk at a Wavelength of 21 cm. I—A New Highly Directional Aerial System', *Australian Journal of Physics*, 1953, **6** p. 190.

(6) Christiansen, W. N., and Warburton, J. A.: 'The Distribution of Radio Brightness over the Solar Disk at Wavelength of 21 cm. III—The Quiet Sun—Two Dimensional Observations', *ibid.*, 1955, **8**, p. 474.

(7) Mills, B. Y., Little, A. G., Sheridan, K. V., and Sleb O.B.: 'A High-Resolution Radio Telescope for Use a 3·5 Metres,' *Proceedings of the Institute of Radio Engineers*, 1958, **46**, p. 67.

(8) Ryle, M.: 'A New Interferometer and Its Application to the Observation of Weak Stars', *Proceedings of the Royal Society*, A, 1952, **211**, p. 351.

(9) Christiansen, W. N., and Mathewson, D. S.: 'Scanning the Sun with a Highly Directional Array', *Proceedings of the Institute of Radio Engineers*, 1958, **46**, p. 127.

(10) Christiansen, W. N., and Mathewson, D. S.: 'The Origin of the Slowly-Varying Component', *Paris Symposium on Radio Astronomy*, 1958, p. 108 (Stanford University Press, 1959).

(11) Dicke, R. H.: 'The Measurement of Thermal Radiation at Microwave Frequencies', *Review of Scientific Instruments*, 1946, **17**, p. 268.

396.933.23

The Institution of Electrical Engineers
Paper No. 3357 E
Jan. 1961

IE PROBLEM OF IMPROVING THE BRITISH INSTRUMENT LANDING SYSTEM LOCALIZER FOR AUTOMATIC LANDING

By A. N. BERESFORD, B.Sc., and J. D. ASTERAKI, M.A.

(*The paper was first received* 19*th November*, 1959, *and in revised form* 8*th August*, 1960.)

SUMMARY

he general outline of the International Civil Aviation Organization ification for an instrument approach system was largely based on pment developed during the last war. Only minor modifications e been incorporated in the specification since then. The paper es the problem of providing a radio guidance system similar to in use but with an improved accuracy. It is suggested that the uracy must be improved by a factor of 3 if the system is to be used landing and not merely approach.

ome parameters required of an instrument landing system localizer established and a line of attack on the aerial problem is described.

LIST OF SYMBOLS

E = Field strength set up by the transmitting aerial.
n = Depth of modulation.
p = Coefficients of θ.
V = Voltage at aircraft receiver input terminals.
θ = Azimuth angle measured from the extended runway centre-line.
ϕ = Angle of the maximum of the tone radiation pattern.
α = Phase angle between carrier voltages radiated by the two aerials.

uffix 1 refers to the tone on the left-hand of an approaching raft and suffix 2 to the tone on the right.

(1) INTRODUCTION

he British Instrument Landing System (B.I.L.S.) consists of e components: the localizer for azimuth guidance, the glide h for height guidance and the marker beacons which give ne indication of the distance to touch down. The purpose of paper is to discuss the performance of the localizer, and so two other components will not be mentioned further.

t is thought, now that several accurate navigational aids are ilable, that the International Civil Aviation Organization .A.O.) coverage requirements should be rewritten to require dance from the B.I.L.S. only in a narrow funnel about the rse line, guidance into the funnel being obtained by other ans. The reasons for this suggestion and the benefits likely esult are discussed.

(2.1) General Description of the Localizer

wo aerials are used to produce radiation patterns similar to se in Fig. 1. A carrier in the frequency band 108–112 Mc/s, plitude modulated at 90 c/s, is fed to one aerial, and the same ier frequency, modulated at 150 c/s, is fed to the other. ateral deviation from the course line is indicated to the pilot a left/right meter actuated by the predominance of one tone he other.

(2.2) I.C.A.O. Requirements

he relevant portions of the I.C.A.O. standards and recommended practices are as follows:

Coverage.—The localizer shall provide signals sufficient to allow atisfactory operation of a typical aircraft installation from the approach end of the runway to a point 600 m above and at least 25 nautical miles from the approach end over a sector 20° wide, centred on the course line up to 7° in elevation above the horizontal measured from the localizer, . . .

The authors consider that this part of the requirement is all

Fig. 1.—Ideal radiated tone patterns.
150 c/s tone predominates on the left.
The curves are of the form $mE \cos(\theta \pm \phi)$.

that is needed and that the remainder of the paragraph which continues as follows could well be deleted:

. . . also to at least 17 nautical miles at a height above the approach end of 600 m in all other directions.

Sensitivity.—The localizer equipment shall be so designed and adjusted that the difference in the depth of modulation of the two tones is 15·5% for an angular deviation of about 2·5° from the course line, and the difference in the depth of modulation shall increase in a substantially linear manner from the course line (where it is zero) to an angular deviation of ±4°. The modulation depth of each tone on course shall be as near as is practicable 20%.

(2.3) The British Instrument Landing System Equipment

The existing aerial array consists of two horizontal half-wave dipoles symmetrically placed on either side of the focus of a parabolic reflector. The reflector is formed of horizontal wires fixed to insulating supports, and it is usually placed at the opposite end of the runway to touch-down, either in line or slightly to one side.

The transmitter is of conventional design. The output of a crystal oscillator is frequency multiplied by 18 to give the carrier frequency. This is fed to two plate modulators, one to give the carrier and 90 c/s sideband frequencies and the other to give carrier and 150 c/s sideband frequencies.

The ground installation is completed by a monitor aerial placed about 400 ft in front of the parabola on the course line. The equi-signal line is made to pass through this aerial by using the signals received, first as a guide while setting up and then to servo-control the depth of modulation produced in the two output units. Interlocks are provided so that, in the event of

ritten contributions on papers published without being read at meetings are ed for consideration with a view to publication.
r. Beresford and Mr. Asteraki are at the Royal Aircraft Establishment.

some parameter altering beyond the range of control, the transmitter is switched off and a standby transmitter is automatically brought into operation.

Provision is made for the transmitter to radiate an identification signal at regular intervals and a voice communication channel is available if required.

(2.4) Criticisms of the Localizer

The main reason why the existing localizer is not completely adequate for automatic landing is because of the low signal/noise ratio on the approach path; the term being used here as the ratio of the desirable information (the signal) to all disturbing sources (the noise). The noise is of two sorts, (i) random fluctuations of voltage at the receiver input terminals (conventional noise), and (ii) signals originating from the localizer but giving false information. Because of the large signal output from the transmitters the first form is not important; the second form of noise can be subdivided into (*a*) signals of the correct (horizontal) polarization reaching the course line after re-radiation from buildings and other objects in the illuminated zone, and (*b*) signals of the wrong (vertical) polarization received by both direct and indirect paths from the aerial. The direct vertically-polarized signal would be of little consequence if the pattern of radiation were the same as that for horizontal polarization. Unfortunately this is not so with the existing aerial.

In an attempt to meet the I.C.A.O. coverage requirement the aerial was made to illuminate a zone about 140° wide. This appreciably straightened the course over that given by an all-round-looking aerial on an average airfield. But even in the restricted zone of 140° there will, on the majority of airfields, be many excellent re-radiators (hangars, buildings, vehicles, aircraft, etc.) in the high-field-strength zone of the aerial. As it is both impracticable and uneconomic to keep such a large area clear, the only solution is to restrict further the radiation from the aerial if a still straighter course is required.

(2.5) Required Localizer Parameters

It is advisable first to establish values for the various parameters of the system. The sensitivity of the off-course indication is controlled by two independent variables: the angle of cut between the radiation patterns and the depth of modulation of the transmitters. For other reasons the depth of modulation of each tone on course is set at 20%. (This means in the British system that each transmitter must be modulated to a depth of 40%.) The angle of cut is thus left as the only variable.

Let the radiation patterns shown in Fig. 1 be $E_1 \cos n(\theta - \phi)$ for the 90 c/s tone and $E_2 \cos n(\theta + \phi)$ for the 150 c/s tone. It is shown in Section 9.1 that the difference in depth of modulation (d.d.m.) off-course given by a perfect system where $m_1 = m_2$, $E_1 = E_2$ and $\phi_1 = \phi_2$ is

$$m \tan n\theta \tan n\phi \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where m is the modulation depth of each transmitter and is 40%.

The d.d.m. will vary linearly with θ only for $n\theta < 30°$. Thus if we require linearity to within 4°, n cannot be much greater than about 7·5 and we have a limit to the narrowness of the patterns. Inserting the value of 15·5% for the d.d.m. when $\theta = 2{\cdot}5°$, we have $\phi = 6{\cdot}54°$.

It is further shown in Section 9·1 that the d.d.m. on the course line (ideally zero) measured by an aircraft when the previous equalities do not hold is

$$\frac{m_1 V_1 \cos n\phi_1 - m_2 V_2 \cos n\phi_2}{V_1 \cos n\phi_1 + V_2 \cos n\phi_2} \quad . \quad . \quad . \quad (2)$$

It has been suggested that the radio equi-signal line should b kept within a sector 8′ wide centred on the extended runwa centre-line. The present I.C.A.O. specification allows a secto 40′ wide. To keep within the 8′ sector, the modulus of th d.d.m. must be less than 0·4% on the extended runway centre line ($\theta = 0$).

The zero d.d.m. line may differ from the $\theta = 0$ line for tw reasons:

(*a*) The transmitter may be incorrectly adjusted.
(*b*) The aircraft may be receiving signals of (i) the correct polariza tion other than by the direct route, and (ii) signals of the wron polarization by both direct and indirect routes.

If we allocate half the allowable error to (*a*) and one-quart each to (*b*)(i) and (*b*)(ii) we see that, in the absence of automat beam correction, the radiated carrier field strengths must b equal to about ½% and the modulation depths must be withi about 0·2% of their nominal value. Also, the received to voltages should be within about ½% of their ideal values a the vertically polarized signal should be about 44 dB down c the horizontally polarized signal.

Tests which have been carried out on the transmitter ha suggest that the stability of the output power and modul tion depth is about two or three times worse than that requi for automatic landing. Some form of automatic beam corr tion is therefore necessary. In the form of correction chos the modulation depths of the two transmitters are altered compensate both for output-power changes and modulati depth changes. Although this works well in practice there a objections. Once the output powers are different the sensitivi is asymmetrical about the course line, and also it has chang on both sides from the required value. This may be unacce able if the localizer is used to obtain rate of change of error well as the error itself.

(2.6) B.I.L.S. with Modified Transmitter

It has been suggested earlier that, with the existing tran mitters, radiation patterns narrower than about $\cos 7{\cdot}5\theta$ w lead to an unacceptable sensitivity condition. It is therefo interesting to compare the B.I.L.S. with a system where t carrier and sidebands can be radiated separately.

Let the radiation patterns for the sidebands be $\cos n(\theta \pm$ and let the carrier radiation pattern be $\cos p\theta$, where p is som what larger than $\frac{1}{2}n$. In Section 9·2 it is shown that the variati of d.d.m. with θ is given by

$$\frac{2 \sin n\theta \tan n\theta}{5 \cos p\theta}$$

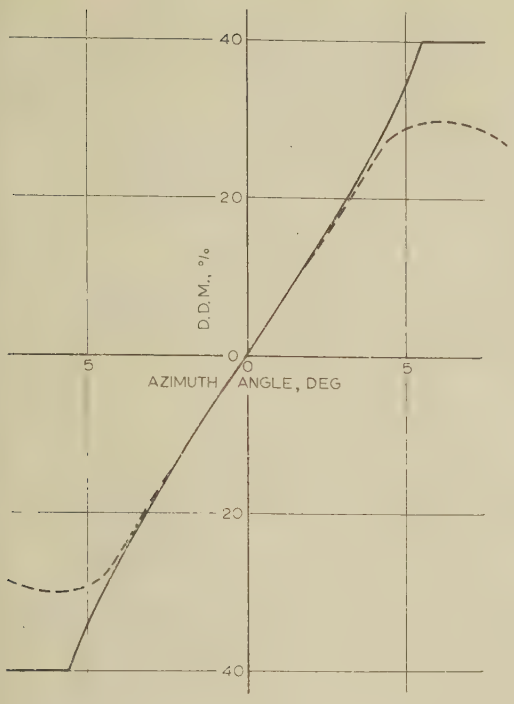
This is shown graphically in Fig. 2 for $n = 12$, $p = 8$ a $\phi = 3{\cdot}0°$. Superimposed on this is a similar curve for t B.I.L.S. with the radiation patterns $\cos 7{\cdot}5(\theta \pm \phi)$. The var tion of d.d.m. with θ is slightly more linear in the region ± than for the B.I.L.S., and as the radiation is more restricted should lead to a straighter course line.

(3) CHOICE OF FREQUENCY FOR AN INSTRUMENT LANDING SYSTEM

All that has been said previously applies whatever the f quency band or plane of polarization chosen. Some of points affecting the choice will therefore be examined.

(3.1) Ground Aerials

One source of temporary beam bends not yet mentioned caused by aircraft flying in the illuminated zone. Televis viewers living in fringe areas are well aware that an aircr



ig. 2.—Theoretical curves showing the variation of d.d.m. with azimuth angle.

—— B.I.L.S. with new aerial.
---- Possible curve after modifying transmitter.

ıg at 1 000 ft or so is in a considerably higher field strength n their receiving aerials, and what is more to the point, the adiated signal from the aircraft may be larger than the signal eived by the direct path. The same is true for the instrument ding system. Recent measurements made by the Blind ıding Experimental Unit have shown that an aircraft flying a few thousand feet can cause the radio equi-signal line to illate about the mean with an amplitude of several degrees. s would be disastrous for an aircraft using the system to l automatically. The only cure is to prohibit flying in that t of the illuminated zone which will give trouble. In order t the difficulties of air traffic control shall not be further eased, this zone must be kept as small as possible and ferably not greater than the 20° azimuthal sector and the 7° levation mentioned in Section 2.2.

he allowable dimensions for the aerial are almost unlimited th and depth, but because the aerial will be an obstruction lying it should not be more than about 12 ft high. Whether h.f. localizer can be used for automatic landing without the stance of another aid, such as leader cables, depends entirely producing an aerial to the above specification. If this cannot done, it seems likely that an auxiliary aid will be required for final guidance to touch-down. It would be relatively easy neet the specification with a microwave system, and this is haps the strongest argument for moving to this frequency d.

(3.2) Equipment

. complete discussion of the merits of each frequency band as as the equipment is concerned is beyond the scope of the er. However, while airfields are close to each other a ıber of B.I.L.S. channels is required, thus requiring the same uency stability and selectivity as in the existing equipment. It is a moot point whether the transmitter and receiver equipment would be simpler at higher frequencies.

For the aircraft aerial, it may be easier to get uniform cover over a 90° sector at the lower frequencies.

(3.3) Other Considerations

Two other points should be considered. The first is usually made by the proponents of a microwave instrument landing system. The argument is as follows: reflections of radio energy from objects in the zone illuminated by the localizer will inevitably cause the radio course line to deviate from the extended runway centre-line in a manner similar to that of Fig. 3. The wavelength of the oscillation is proportional to

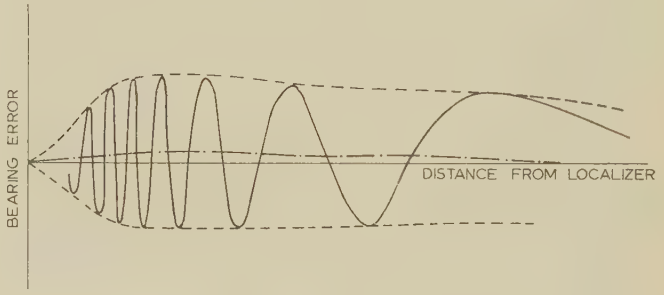


Fig. 3.—Calculated error angle for a typical case.

—— Error.
---- Envelope of error.
—·— Mean of envelope.

the radio wavelength, and hence, if we make this small enough, the wavelength of the deviations will be so small that the aircraft will not be able to follow them and hence will fly on a smooth path—the mean of the envelope. There are two points to be made against this argument: (*a*) There are always positions for re-radiating objects which will make the wavelength of the deviations large enough for an aircraft to follow whatever the radio wavelength, and (*b*) where the deviation wavelength is small, the mean of the envelope has a lateral displacement from the extended runway centre-line. Reference 1 provides further information on this point.

The second point was brought to the authors' attention by Mr. J. S. Shayler. He suggested that the position of the effective

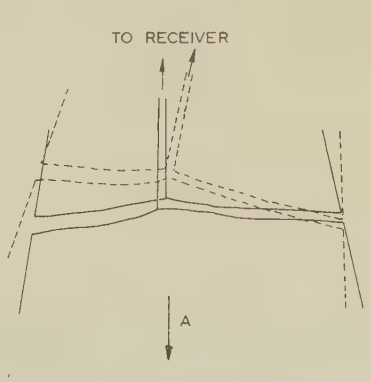


Fig. 4.—Aerial system which would suffer from apparent course wandering.

centre of the aircraft aerial may vary with the aspect of the aircraft. This may best be explained by the following example. If the receiving aerial consists of two dipoles inclined at a small angle (Fig. 4), radiation from a transmitter in the direction A

will be received equally. Hence if the transmitter were defining an equi-signal line, a null would be indicated when the geometric centre of the aerials was on this line. If the receiving system is now rotated through a small angle to the position indicated by the dotted lines, almost all the signal will be received by the left-hand dipole. The aerials will have to be moved laterally by almost half their spacing to restore the null indication.

This kind of situation may be present in all aircraft aerial installations, and it could lead to apparent course-line wandering that is not negligible. This is an argument in favour of an aircraft aerial which is completely independent of the aircraft structure, and it may be easier at high frequencies.

There are two other causes of apparent course shift with aircraft heading. The first occurs where there is a beam bend. If the aircraft aerial radiation pattern is not truly spherical but has holes in it, then, for certain aircraft headings, the re-radiating object causing the beam bend may be in the same direction from the aircraft as a hole in the pattern. The meter zero will then be the true equi-signal line. However, if the aircraft were to yaw slightly so that the hole direction was no longer towards the re-radiator, the meter zero would be on the bend. The magnitude and sensitivity of the shift with heading will, of course, depend on both the size of the beam bend and the steepness of the sides of the hole. The second cause is similar, the difference being that the hole in the radiation pattern is now to signals of vertical polarization, and these signals are propagated directly from the localizer aerial.

(4) CHOICE OF POLARIZATION

If we remain in the 3 m band there may be some advantage in changing the polarization. The work of the Radio Research Station[2,3,4] at Slough on propagation of this band at grazing incidence has shown that, when the transmitting and receiving aerials are more than a few feet above flat ground, the two

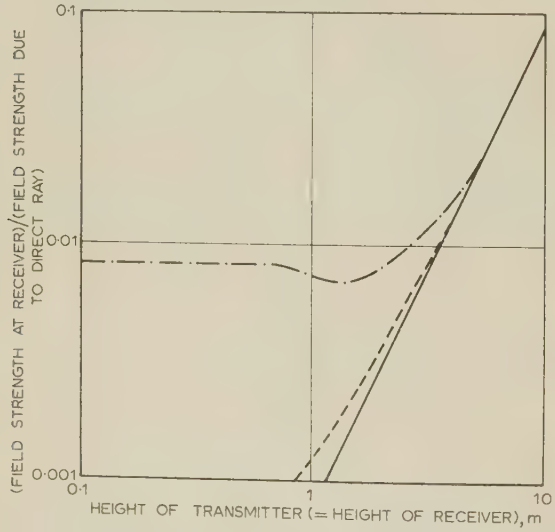


Fig. 5.—Difference in field strength close to the ground between vertical and horizontal polarizations.

Distance between transmitter and receiver = 4·5 km.
$K = 5$ $\sigma = 2 \times 10^{-8}$ e.s.u.
——— Reflection coefficient = −1 ---- Horizontal polarization.
—·— Vertical polarization.

planes of polarization are propagated almost identically, but very close to the ground the vertically polarized signal is stronger. Fig. 5 has been calculated from the formulae given in Reference 2. It is clear that the vertical signal will be affected by low obstacles

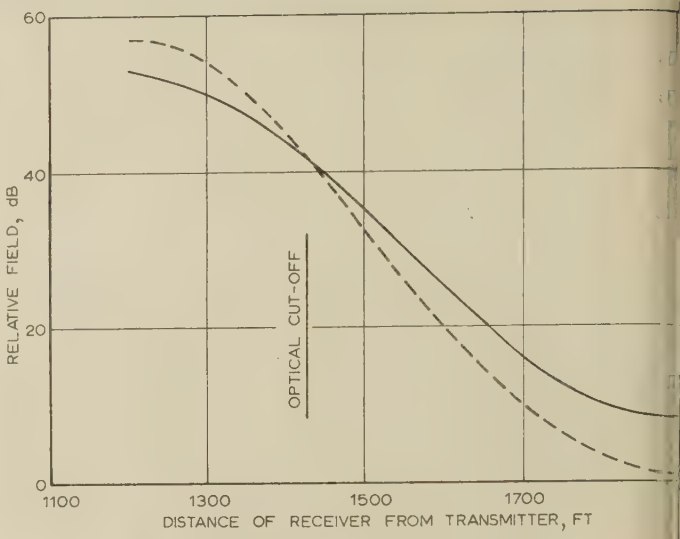


Fig. 6.—Received signal at Farthing Down.

$\lambda = 2 \cdot 25$ m
Vertical-plane contour is very nearly an arc of a circle of 1 250 ft radius.
——— Vertical polarization.
---- Horizontal polarization.

such as runway lights, grass, small shrubs, radar reflectors, etc to a much greater extent than a horizontally polarized sig On this basis horizontal polarization is preferable but it has disadvantage. Reference 4 describes some measurements ma as a receiver was moved over the brow of a hill into the opti shadow, and Fig. 6, which is taken from that paper, shows th when the transmitter and receiver are in line of sight the ho zontally polarized signal is slightly stronger, but, in the shado the vertically polarized signal is propagated preferentially. is therefore essential to ensure that, if horizontal polarization used for a localizer, it should be optically visible from an aircra during all phases of the approach. In this case, horizont polarization is preferable.

(5) AERIAL SPECIFICATION

It is now possible to prepare a specification for an aerial f use with the existing B.I.L.S. transmitters:

(*a*) The tone radiation patterns should be very near cos 7·5(θ + cos 7·5($\theta - \phi$), where ϕ may be varied between 5·7° and 7·4° accommodate different localizer aerial-to-touch-down distances).
(*b*) Outside the angles ±25° the radiation should be as small possible, the target being 52 dB down on the on-course signal.
(*c*) The vertically polarized signal should everywhere be as sm as possible, the target being 40 dB down on the wanted on-cou signal.
(*d*) The radiation should as far as possible be kept below 7° elevation, the target above this being 52 dB down on the on-cou signal.

An aerial which very nearly meets clauses (*a*), (*b*) and (*c*) this specification has been erected. Although there are so unexplained phenomena associated with it, in general the te show that the course line is straighter than that given by parabolic reflector.

The new aerial consists of two electromagnetic horns, ea having an aperture width of 27 ft, a depth of 80 ft and a hei of 8 ft. The method of construction was to erect a frame tubular scaffolding and then to fasten ½ in mesh wire netting the inside of the frame. A tolerance of ±3 in was placed the straightness of the sides. The horns are fed by reson slots backed by cavities, an H_{01} mode being excited in them

uitable probe. A small loop and crystal detector provide a venient monitor of the cavity field strength.

he only other change which has been made in the equipment o replace the r.f. cables feeding the aerials. A standard allation uses U.R.75 cable. It was considered that the eral stability of this cable was not sufficient for the accuracy uired, and so a large-diameter cable with a polythene helical nbrane was substituted.

(6) CONCLUSIONS

ome of the parameters required from a localizer have been blished, and the first attempt at an aerial to give a better ormance has been described. Although on balance it seems a microwave system might be more suitable for automatic ling, in view of the capital invested in the existing equipment ould be folly to abandon it without exploring every possi- y for improving the performance sufficiently for it to be l for automatic-landing without the aid of a system such as er cables.

(7) ACKNOWLEDGMENTS

he authors would like to thank Dr. J. S. McPetrie, Mr. W. J. rnley, Mr. J. S. Shayler, Dr. K. E. Bucks and Mr. N. H. fle for their help and advice during many discussions, and Messrs. B. H. Bryan, J. P. Bentick, P. J. Barley and the other abers of the staff at the Royal Aircraft Establishment who ed in the construction of the new aerial. They would also to thank Dr. A. G. Touch, who has succeeded Dr. J. S. Petrie as Head of Radio Department at the Royal Aircraft blishment since this paper was started, for the simulating ussions they have had with him.

cknowledgment is also made to the Controller, H.M. ionery Office, for permission to publish the paper.

(8) REFERENCES

JOHNSON, W. A., BERESFORD, A. N., and ROTHE, P.: 'A Method of Studying Track Errors in C.W. Landing Aids caused by Scattered Radiation', Unpublished Ministry of Supply report.

MCPETRIE, J. S., and STICKLAND, Miss A. C.: 'Reflection Curves and Propagation Characteristics of Radio Waves along the Earth's Surface', *Journal I.E.E.*, 1940, **87**, p. 135.

MCPETRIE, J. S., and SAXTON, J. A.: 'An Experimental Investigation of the Propagation of Radiation having Wavelengths of 2 and 3 Metres', *ibid.*, 1940, **87**, p. 146.

MCPETRIE, J. S., and FORD, L. H.: 'An Experimental Investigation on the Propagation of Radio Waves over Bare Ridges in the Wavelength Range 10 cm to 10 m', *ibid.*, 1946, **93**, Part IIIA, p. 527.

(9) APPENDICES

(9.1) Derivation of the Off-Course Difference in Depth of Modulation for the B.I.L.S.

Consider the radiation patterns of Fig. 1. The 90 c/s tone pattern is $m_1E_1 \cos n(\theta - \phi_1)$ and the 150 c/s pattern is $m_2E_2 \cos n(\theta + \phi_2)$, and therefore the difference in depth of modulation (d.d.m.) is

$$\frac{E_{90} - E_{150}}{E_{carrier}}$$

where E is the radiated field strength.

Therefore the d.d.m. is

$$\frac{m_1E_1 \cos n(\theta - \phi_1) - m_2E_2 \cos n(\theta + \phi_2)}{[E_1 \cos n(\theta - \phi_1) + E_2 \cos n(\theta + \phi_2)] \cos \frac{\alpha}{2} + j[E_1 \cos n(\theta - \phi_1) - E_2 \cos n(\theta + \phi_2)] \sin \left(\frac{\alpha}{2}\right)}$$

for $$-\left(\frac{90}{n} - \phi_1\right) \leqslant \theta \leqslant \left(\frac{90}{n} - \phi_2\right)$$ (i.e. θ between δ_1 and δ_2 of Fig. 1)

A series of tests has shown that, once the transmitter has been correctly adjusted, $\alpha < 1°$. Thus for all practical purposes the d.d.m. may be expressed as

$$\frac{m_1E_1 \cos n(\theta - \phi_1) - m_2E_2 \cos n(\theta + \phi_2)}{E_1 \cos n(\theta - \phi_1) + E_2 \cos n(\theta + \phi_2)} = \frac{\cos n\theta(E_1m_1 \cos n\phi_1 - E_2m_2 \cos n\phi_2) + \sin n\theta(E_1m_1 \sin n\phi_1 + E_2m_2 \sin n\phi_2)}{\cos n\theta(E_1 \cos n\phi_1 + E_2 \cos n\phi_2) + \sin n\theta(E_1 \sin n\phi_1 - E_2 \sin n\phi_2)}$$

If the transmitter is so adjusted that

$$E_1 = E_2 = E$$

$$m_1 = m_2 = m$$

$$\phi_1 = \phi_2 = \phi$$

then the d.d.m. is $m \tan n\theta \tan n\phi$ . . . . . (3)

This will be 10% greater than a linear function of θ for $n\theta = 30°$. Thus, if we require linearity to within 10% to $\theta = \pm 4°$, n cannot be greater than 7·5. In part of the I.C.A.O. specification, the sensitivity is specified as a d.d.m. of 15·5% for a stated lateral deviation from the course line at a certain distance from touch-down. To accommodate variations in the distance between touch-down and the localizer aerial the angle at which this occurs varies between 2° and 3°. The sensitivity is given the correct value by altering ϕ, as m has been fixed for other reasons. ϕ lies between 5·7° and 7·4°.

If $E_1 \neq E_2$, $m_1 \neq m_2$, $\phi_1 \neq \phi_2$, but we are on the line $\theta = 0$.

$$\text{D.D.M.}_{(\theta = 0)} = \frac{E_1m_1 \cos n\phi_1 - E_2m_2 \cos n\phi_2}{E_1 \cos n\phi_1 + E_2 \cos n\phi_2}$$

In general, the aerial is a mechanically rigid structure and ϕ is almost invariable. In the paper it is suggested that the radio equi-signal line should be kept within $\pm 4'$. This means that the modulus of d.d.m. for $\theta = 0$ must be less than about 0·4%.

(9.2) Derivation of the Off-Course D.D.M. for a Separated Sideband

Consider the position when the sidebands and carrier are radiated separately, the sidebands in $\cos n(\theta \pm \phi)$ patterns as

in Fig. 1, and the carrier in $\cos p\theta$ pattern. Assuming equality between E_1, E_2, etc., the variation of d.d.m. with θ is given by

$$\frac{\cos n(\theta - \phi) - \cos n(\theta + \phi)}{5 \cos n\phi \cos p\theta}$$

The term $5 \cos n\phi$ represents the on-course depth of modulation. The factor 5 comes in because of the required 20% depth of modulation. Therefore, the d.d.m. is

$$\frac{2}{5} \frac{\sin n\theta}{\cos p\theta} \tan n\phi$$

If we require coverage to $\pm 9°$, say, then n cannot be greate than 12. If we put $n = 12$ and $p = 8$ we have for $\theta = 2{\cdot}5°$

$$15{\cdot}5\% = \frac{2}{5} \frac{0{\cdot}5}{0{\cdot}9397} \tan n\phi \times 100$$

or

$$\phi = 3{\cdot}0°$$

A plot of d.d.m. against θ is shown in Fig. 2. It is slight more linear in the region $\pm 4°$ than the B.I.L.S. with the narrowe beam aerial even though it illuminates a much smaller secto there being no radiation outside $\pm 11\frac{1}{4}°$ compared with abo $\pm 18\frac{1}{4}$ for the narrowest-beam B.I.L.S.

396.677 : 621.397

The Institution of Electrical Engineers
Paper No. 3334 E
Jan. 1961

THE USE OF A HIGH-GAIN TELEVISION TRANSMITTING AERIAL IN A POPULOUS AREA

With particular reference to the Crystal Palace Station

by G. D. MONTEATH, B.Sc., A.Inst.P., Associate Member, G. H. MILLARD, B.Sc., A.Inst.P., and D. J. WHYTHE, B.Sc.(Eng.), Associate Member.

(*The paper was first received* 1*st March, in revised form* 24*th May, and in final form* 7*th July*, 1960.)

SUMMARY

ne paper is concerned with the provision of satisfactory signals ewers in the immediate vicinity of the Crystal Palace television on, where an 8-tier aerial is sited in an urban area. The problem complicated by differences between the halves of the aerial, but a factory performance was ultimately obtained. The only serious epancy between theoretical and experimental results could be oved by taking account of reflection at the ground, ignoring the t of buildings.

order to provide continuity of service in spite of any single kdown, the halves of the aerial were originally supplied from rate transmitters; this arrangement had to be abandoned, since it ed impossible to provide undistorted signals at receiving points e the contributions from the transmitters were in phase opposition.

(1) INTRODUCTION

t a television transmitting station it is advantageous to use a -gain aerial, usually comprising several tiers, in order to ine most of the radiated energy to directions near the hori- al. If the tiers of the aerial are identical and equally spaced, gain is greatest when the radiating currents in all of them co-phased and approximately equal in magnitude. Fig. 1 vs the theoretical vertical radiation pattern (v.r.p.) of a al 8-tier aerial. In practice the minima are not zeros as vn, but they are associated with annuli in which the field ngth is much less than at other points in the vicinity. The strength in these annuli will rarely, if ever, be insufficient satisfactory reception, but the signals are likely to be dis- ed for two reasons: first, the v.r.p. will usually vary with uency, so that wide disparities can exist between the levels ifferent components of the radiated spectrum; secondly, the imity of areas where the field strength is much greater than ne receiving point increases the likelihood of ghost images g caused by multi-path propagation. The area in which e difficulties are encountered increases with both the gain the height of the aerial; for a typical 8-tier aerial the radius he outermost annulus of minimum field is equal to about t times the aerial height.

is always possible to modify the v.r.p. of the transmitting al so as to fill in the minima (this procedure may be termed filling'), but the gain of the aerial is thereby reduced. In r to avoid intolerable distortion for a few viewers living to the station, many viewers, including those on the fringe he service areas, must suffer a small loss of field strength. best compromise can be made only as a result of experience. ne U.S.A., where stations are usually sited in towns, the need gap filling became apparent soon after television broad- ing began, and a number of methods were devised.[1] The blem was not encountered so soon in the United Kingdom, where high-power stations are usually sited in the country, so that few viewers live very near to them. The Crystal Palace station[2] was exceptional in this respect, having an 8-tier aerial and being sited in an urban area; the paper records the experience gained at this station.

At Crystal Palace circumstances forced the design of an aerial with dissimilar halves, so that the v.r.p. varied with azimuth and gap filling was made more difficult. Although such an arrangement would not have been adopted without good reason, it is useful to record the extent to which the resulting difficulties were overcome. In populous areas a tapering tower may be the only type of mast permitted, and the mechanical design might be simplified in some cases by not insisting on a long uniform section for supporting the aerial.

(2) SOME METHODS OF GAP FILLING

The only methods of gap filling considered in the paper amount to modifying the amplitude and phase distribution of the radiating currents. The required amplitudes and phases can be determined in either of two ways:

(*a*) The aerial is regarded as an aperture over which the current distribution is determined in terms of the Fourier transform of the required radiation pattern,[3,4] this method leads to a different radiating current in each tier.

(*b*) The v.r.p. obtained with all tiers fed equally and in phase is taken as a starting point and the simplest changes are made to provide the degree of gap filling required.

Of these, (*a*) is probably the more suitable for use in Bands IV* or V, where a large number of tiers may be used and where the modification of the v.r.p. must be considerable. Method (*b*) is preferable in Band I, where it is unusual for more than eight tiers to be employed. Either method might be used in Band III. In what follows, only method (*b*) will be considered.

It is convenient to define the phase associated with the v.r.p. with reference to the phase of the field which would be radiated by an isotropic source at the centre of the aerial. Applying this convention to the v.r.p. shown in Fig. 1, the phase is constant over each lobe, adjacent lobes being in antiphase. It follows that the minima could be filled in by installing an additional tier at the centre of the aerial and feeding it in phase quadrature to the other tiers. This method would, however, be inefficient, in addition to requiring an additional tier, since the central tier would radiate half its energy upwards. A better method would be to modify the amplitudes and phases of the radiating currents in the two central tiers of the original aerial by superimposing upon them additional components of current. These would be so chosen that in isolation their v.r.p. would be tilted downwards and would be in phase quadrature to the primary v.r.p. Another method is to feed all tiers of the aerials in phase but with unequal amplitudes. If the amplitudes are tapered from one end of the

itten contributions on papers published without being read at meetings are d for consideration with a view to publication.
e authors are with the British Broadcasting Corporation.

* The frequency limits of these bands are: Band I, 41–68 Mc/s; Band III, 174–216 Mc/s; Band IV, 470–585 Mc/s; Band V, 610–960 Mc/s.

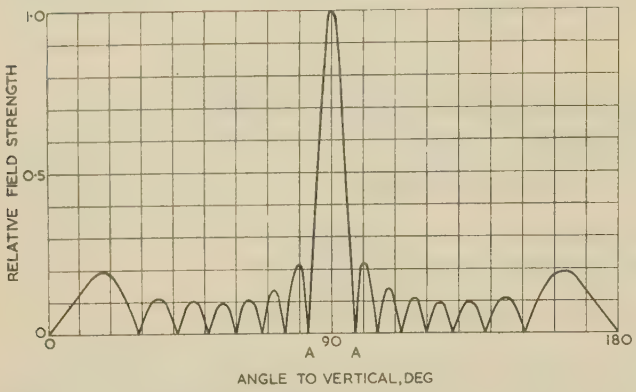


Fig. 1.—Vertical radiation pattern of a typical 8-tier aerial: all tiers fed with equal co-phased currents.

aerial to the other, all the minima are filled in to some extent; this method is quite efficient but rather complicated. By feeding the halves of the aerial with unequal powers alternate minima may be filled,[5] including the minimum nearest to the main lobe (A in Fig. 1), which is usually the most important since it affects the largest area.

(3) THE EFFECTIVE VERTICAL RADIATION PATTERN AT FINITE DISTANCE

Up to this point the problem has been considered in terms of the v.r.p., i.e. the free-space field at a great distance as a function of the zenithal angle, but two modifications to this simple representation must be considered: first, the distances involved in practice may not be sufficient to be regarded as infinite (unfortunately this will usually be the case in Band I, although seldom in Band III and probably never in Bands IV and V); secondly, the effect of reflection at the ground cannot be ignored.

One way in which finite distance modifies a v.r.p. such as that shown in Fig. 1 is illustrated in Fig. 2. Suppose that P is a

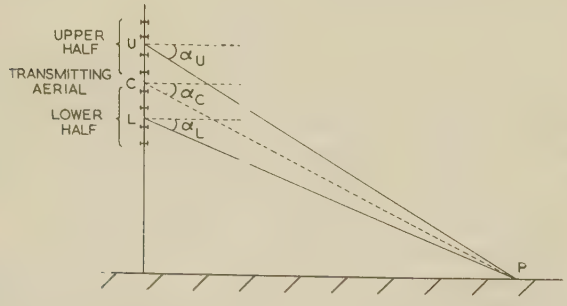


Fig. 2.—The effect of finite distance on the relative amplitudes and phases of the contributions from the halves of the aerial.

point corresponding to minimum A, at which the contributions from the two halves of the aerial are in antiphase. It will be seen that the angle α_U is greater than α_L; in other words, the signal from the upper half of the aerial is received from a greater angle to the maximum of the v.r.p. than that from the lower half. It follows that the contribution to the field strength from the upper half will be weaker than that from the lower half. This effect will be accentuated slightly by the fact that UP is greater than LP. Since the contributions are unequal in magnitude, they cannot cancel one another completely and the minimum in the v.r.p. will be partially filled in. The discontinuous change of phase by 180° between adjacent lobes is, of course, converted to a continuous change.

The argument outlined above may be extended to show tha all the minima will be partly filled. These effects depend on th physical length of the aerial in relation to its height abov ground, and it is for this reason that their importance is mainl confined to Band I.

At first sight it might appear that the effect of finite distanc must be wholly beneficial, since it provides a measure of ga filling. But if this proves insufficient, the provision of additiona gap filling by modifying the radiating currents becomes mor difficult.

(4) THE EFFECT OF REFLECTION AT THE GROUND

If the ground were not encumbered with buildings and if th height of the receiving aerial were known, the effect of grou reflection could be determined readily. For the frequencies a angles of incidence with which we are concerned, it would b reasonable to assume a reflection coefficient of -1.

For the Crystal Palace aerial, most of the receiving aeria affected by minima in the v.r.p. were below the height giving th greatest received signal. In this region the effect of reflection to enhance the relative contributions made to the received fi by the upper parts of the transmitting aerial. This eff opposes that discussed in Section 3, so that the two effects to to cancel one another.

When the Crystal Palace aerial was being designed it wa decided to ignore ground reflection, in view of the uncertai effect of the heavy concentration of buildings. It now appe that this decision was incorrect.

(5) THE CRYSTAL PALACE AERIAL

The television transmitting station at Crystal Palace[2] w planned when television was being extended to higher frequen bands and the possibility of additional services was under co sideration. It was therefore necessary to make provision f these bands before the requirements for them had crystallize As a result, changes had to be made in the design of the Band aerial when it was too late to redesign it *ab initio*. A bri account of these changes is necessary to explain its present form

The site was restricted in area, necessitating the use of a se supporting tower instead of the more usual stayed mast. It w originally planned that the tapering tower should be surmount by two parallel-sided sections carrying an 8-tier Band I aer and an 8-tier Band III aerial, the latter to be capable of radiati two programmes. It was subsequently decided that the tow must also be capable of supporting an aerial for Band IV or which would project from the top. This change made it nec sary to increase the cross-sections of the upper sections of t tower, and it was decided to accept the resulting impairment t the uniformity of the horizontal radiation pattern (h.r.p.) of t Band I aerial.

Finally, it was decided to provide space for two 16-tier Band aerials instead of one 8-tier aerial, since the B.B.C. had agre to provide a separate 16-tier aerial at the top of the tower t the Independent Television Authority. This decision, whi was made while the tower was in course of construction, entail reducing the height of the Band I aerial by mounting its lov four tiers on the tapering support tower, and making the verti spacing of the upper four tiers less than the optimum. The lov four tiers were entirely redesigned, but it was too late to ma any change in the design of the upper four tiers. The final fo of the aerial[6] is shown in Fig. 3.

(6) HORIZONTAL RADIATION PATTERNS

The different forms of the upper and lower halves of the aer result in h.r.p.'s which differ in both amplitude and phase. T

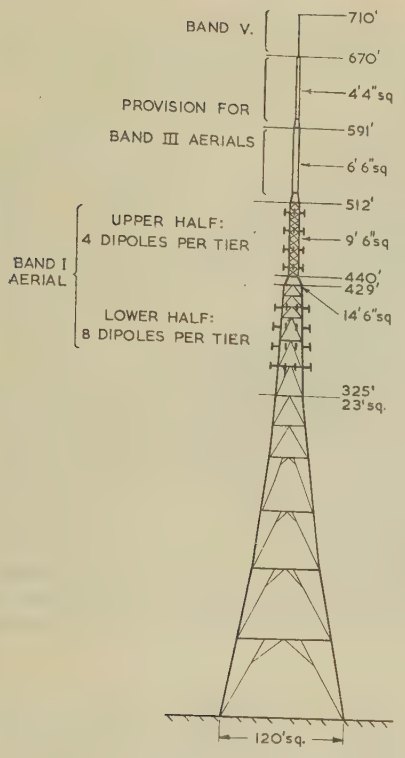


Fig. 3.—The Crystal Palace tower.

of the combined aerial is the vector sum of the contributions from the halves. Furthermore, since each half is fed from parate feeder, the phasing of the inputs to the two halves influence the h.r.p. of the combined aerial.

e h.r.p.'s were deduced from model measurements at oximately one-tenth scale. Two separate models were used, epresenting one of the four tiers on the tapering support r and the other one of the four tiers on the parallel-sided on above. The effect of the tapering form of the tower on h.r.p.'s of the lower four tiers was assessed by slightly ging the frequency (and hence the model scale factor), at ame time making the appropriate changes in the dipole ions. The amplitude and phase of the field from each l was measured relative to that at an arbitrary bearing by l method.[7]

e measured amplitude and phase of the h.r.p. of one tier ur dipoles representing the upper half of the aerial is shown g. 4(*a*). The model of the lower half of the aerial was used termine the arrangement of dipoles which gave the most rm h.r.p., the cross-section of the model mast corresponding at of the full-size tower at the centre of the four tiers. In rrangement chosen, eight dipoles were mounted in pairs, from the mast faces, with a separation of 0·48λ. These ngs were adopted for all four tiers without regard to the , which was found to have little effect on the h.r.p. The ured h.r.p. of one tier of eight dipoles, shown in Fig. 4(*b*), therefore taken to be the resultant pattern of the lower tiers.

e h.r.p. of the two halves combined was deduced by adding wo fields at all bearings with a chosen phase difference. The ing selected gave a compromise between the greatest mean (averaged over the h.r.p.) and the greatest gain at the ma of the h.r.p. It was such that the contributions from alves of the aerial were in phase at bearings of 17° to the of the tower. This condition provided a means of setting

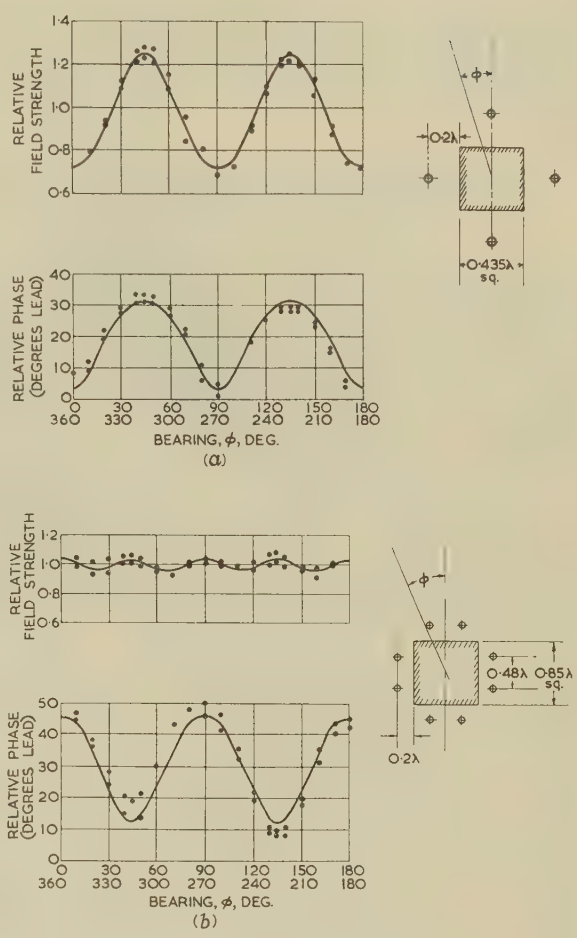


Fig. 4.—Measured h.r.p.'s of model aerial.

● ● ● Measured points.
——— Average over eight octants.
(*a*) Upper four tiers.
(*b*) Lower four tiers.

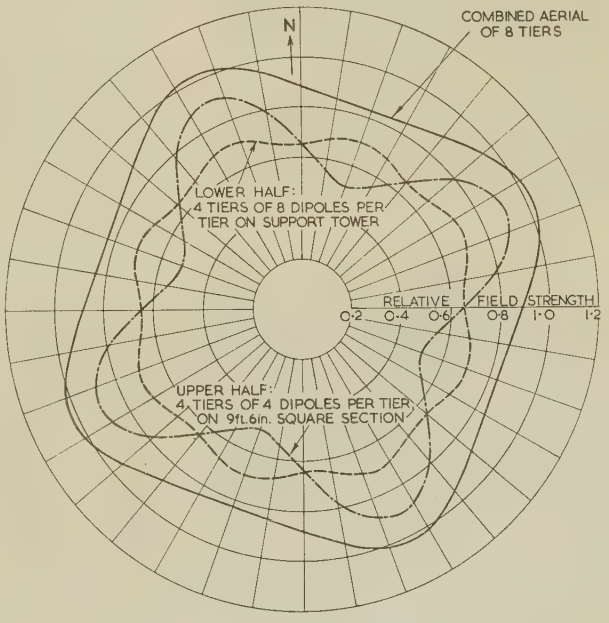


Fig. 5.—H.R.P.'s of the final aerial.

up the full-scale aerial; a receiving point was established on the specified bearing and the relative phasing of the feeds to the halves of the aerial was adjusted for maximum signal. Fig. 5 shows the h.r.p. of the combined aerial and of each half taken separately.

(7) CALCULATION OF FIELD STRENGTH NEAR THE TRANSMITTING STATION

(7.1) Method of Calculation

The unusual form of the aerial complicated the calculation of the field strength variation near the transmitting station but had the incidental advantage of providing a more critical comparison with measurements. In order to get a theoretical result it was necessary to make extensive approximations and to combine theoretical results with experimental data obtained from the scale models.

The basic assumptions were as follows:

(*a*) Values of field were calculated initially for a horizontal plane passing through a point 30 ft above the base of the mast. It was then assumed that variations in ground level would merely displace the field-strength contours radially in proportion to the depth of the surface of the ground below the centre of the aerial.

(*b*) Reflection at the ground was neglected.

(*c*) The v.r.p. of each tier was assumed to vary as $\sin\theta$, where θ is the angle to the vertical of the line joining the transmitting and receiving aerials.

(*d*) The v.r.p. of the receiving aerial was assumed to vary as $\sin\theta$.

(*e*) The amplitudes and phases of the contributions from the eight tiers of the transmitting aerial were adjusted to take account of the amplitude and phase variations of the h.r.p.'s, which were assumed to be independent of θ.

(*f*) The mean phase of each contribution, averaged over the h.r.p., was corrected to take account of the non-uniformity of the tower. This correction, which was different for each of the lower four tiers and for the upper four tiers, was obtained from Carter's formulae[8] for the radiated fields of aerials round cylinders.

Values of the field were calculated for ten distances between 250 and 5000 ft at bearings of 0°, 17°, 30° and 45° to the mast faces.

(7.2) Performance in the Absence of Gap Filling

If reception in the immediate vicinity did not have to be considered, the greatest gain would be achieved by adjusting the phases of the radiating currents in the lower four tiers to bring their contributions to the distant field into phase. This entailed a phase difference of 17°* between the currents in adjacent tiers, the phase in the upper tiers being advanced.

With the phases of the radiating currents controlled in this manner, the field strength near to the ground has been calculated as a function of th distance for various directions, making the assumptions outline in Section 7.1. The broken curves in Fig. 6 show these resul for bearings of 0°, 17°, 30° and 45° to the faces.

Logarithmic scales have been used in Fig. 6 in order facilitate the estimation of the variation of field strength wi frequency at any one point. It may be shown that the effe of a small change in frequency is mainly to move the curv bodily to the right or left, since the range of a particular featu of the curve varies in inverse proportion to the frequency. Th displacement of the curves corresponding to the differen between the carrier frequency (45 Mc/s) and the extreme sid band frequency (42 Mc/s) is indicated on the Figure. The greate variation with frequency occurs near the minima, where t curves are steep. From the upper half of Fig. 7, which indica the regions in which the extreme variation in field strength wi the band 42–45 Mc/s is within given limits, it will be seen th there are considerable areas in which the variation is gre than 10 dB; appreciable distortion would be expected in the areas. As a result of these calculations, it was decided that ga filling would be required.

It is difficult to devise means of achieving sufficient gap fi by using the curves in Fig. 6 only. A better guide was obtai by plotting field strength as a vector on the Argand diag

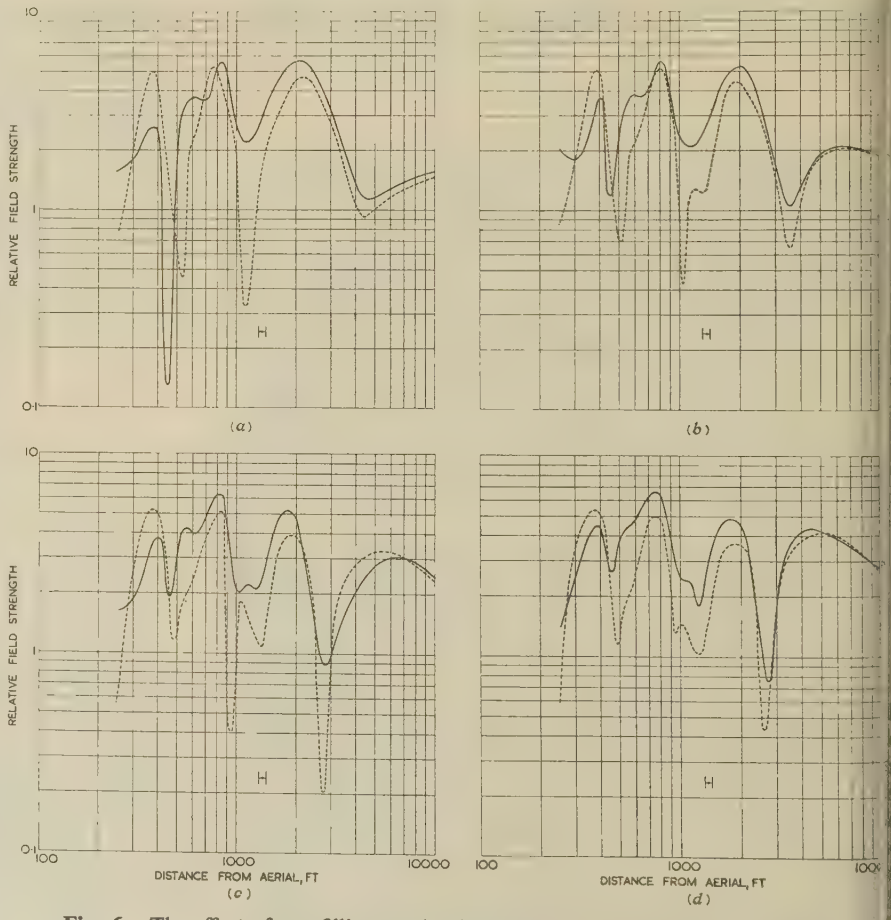


Fig. 6.—The effect of gap filling on the theoretical field strength near the aerial.

---- Field strength without gap filling.
—— Field strength with gap filling.
|←| Horizontal displacement of curves for a frequency change of 3 Mc/s.
(*a*) 0° bearing to tower face.
(*b*) 17° bearing to tower face.
(*c*) 30° bearing to tower face.
(*d*) 45° bearing to tower face.

* The correspondence between this phase difference and the bearing of 17° in Section 6 is accidental.

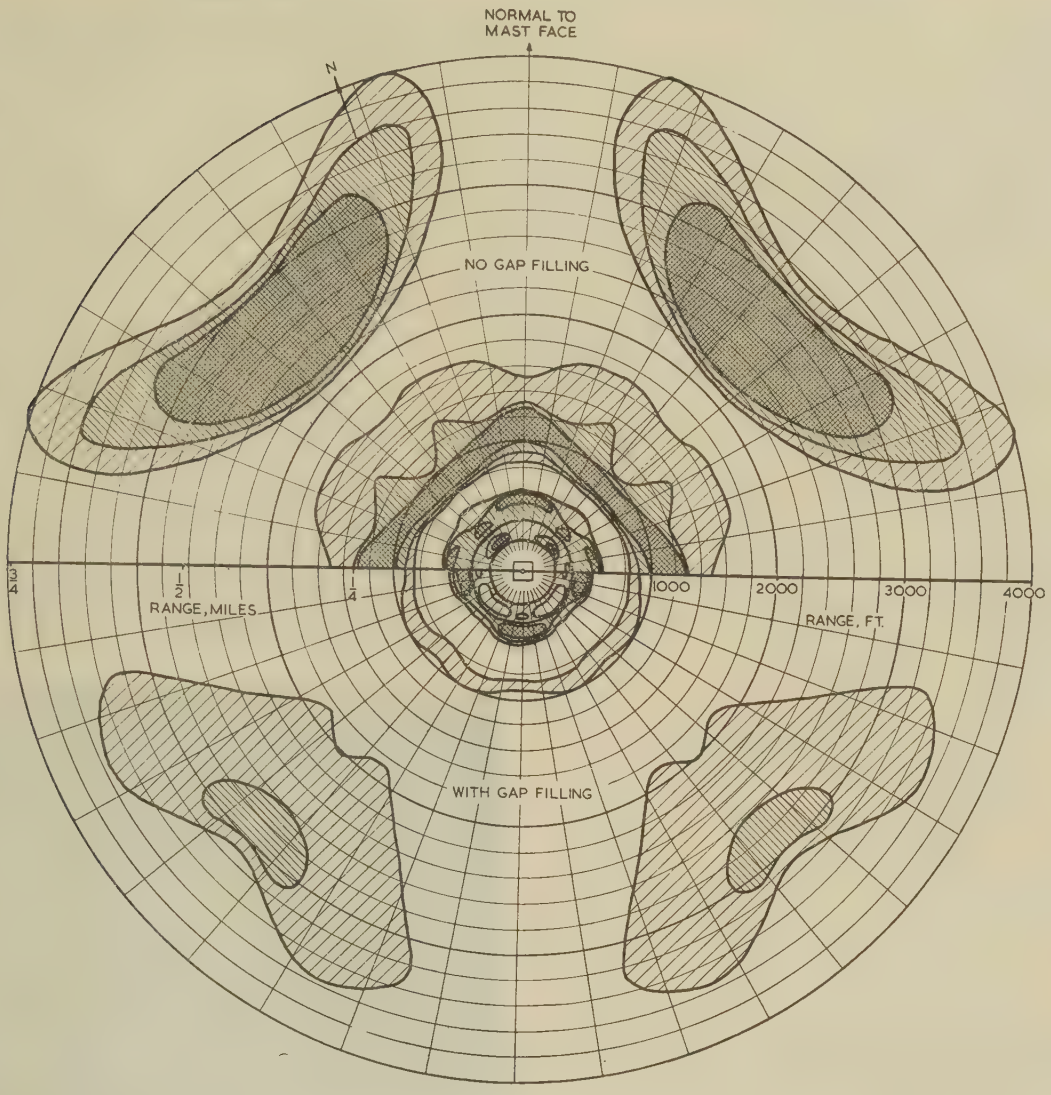


Fig. 7.—Effect of gap filling on reception near the aerial.

Variation of field strength over the band 42–45 Mc/s.

< 3 dB.
> 3 dB.
> 6 dB.
> 10 dB.

ıg as a phase reference a hypothetical source at the centre of aerial system. As an example, Fig. 8(*a*) shows an Argand ram plot for a bearing at 30° to a face. If the effect of finite ınce were negligible and if all the tiers of the aerial were tical and equally spaced, carrying equal radiating currents, v.r.p. would be co-phased. The locus shown in Fig. 8(*a*) ld then be described by a point oscillating on the real axis. much more complex form of the locus in Fig. 8(*a*) is due e effect of finite distance, the unequal spacing between tiers e aerial and the unequal h.r.p.'s of the halves of the aerial. e effects partially fill all the minima, and at the particular ng considered in Fig. 8(*a*) only two minima (having ranges pproximately 1 000 and 3 000ft) appear to require further g.

(7.3) Initial Method of Gap Filling

The method initially adopted was determined by a process of trial and error using the method of calculation outlined above. A change in the radiating currents in one or more tiers was postulated and its effect assessed by regarding the change as due to the superposition of additional components of radiating currents. The field due to these components, which varied with distance, was added as a vector so as to modify the locus of the field on the Argand diagram.

It was necessary to bear certain practical considerations in mind. In the first place it was undesirable to make any change in the feeding arrangements in the upper half of the aerial, since this feeder system was already being manufactured.

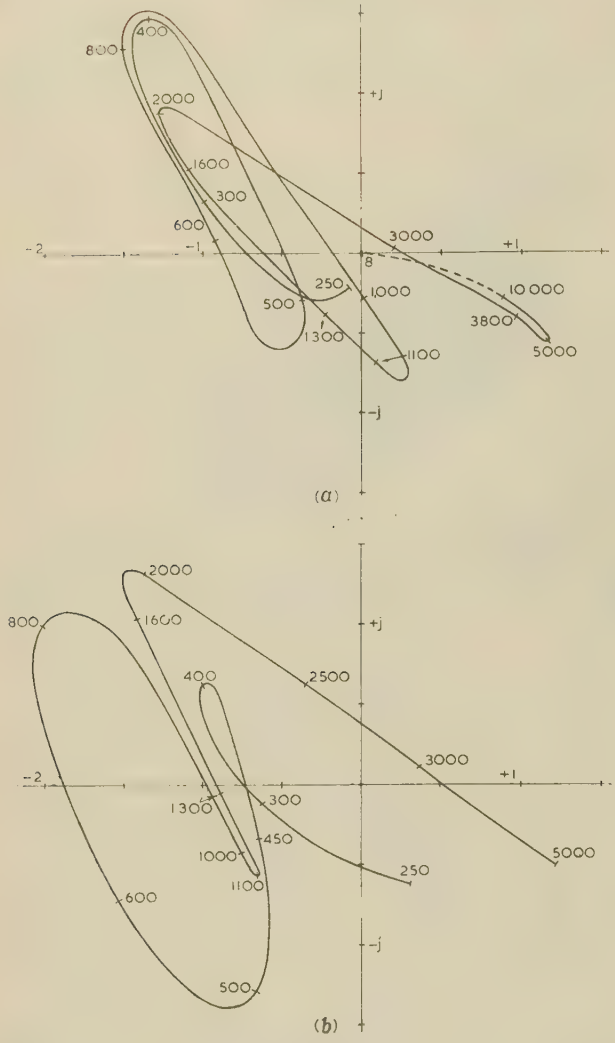


Fig. 8.—Complex relative field strength near aerial, calculated for bearing 30° to tower face.

Numbers on curves refer to distances in feet from base of aerial.

(*a*) Without gap filling.
(*b*) With gap filling.

Secondly, it was undesirable to feed unequal powers to the halves of the aerial, since each half was to be fed independently from a separate transmitter.

With these considerations in mind, the method selected was to advance the phase of the fifth tier (number tiers from 1 to 8 from top to bottom of the aerial system) by 38° and to retard the phase of the sixth, seventh and eighth tiers by 12°. This change left the phase of the resultant field of the lower half of the aerial (tiers 5–8) unchanged. The phases of the radiating currents associated with the resulting contributions to the distant field at bearings of 17° to the faces of the tower are shown in Table 1 and compared with corresponding figures that would be obtained in the absence of gap filling.

It is easy to see the manner in which the field strength in the outermost minimum, which is by far the most important, is increased by making these phase changes. As stated in Section 3, the lower half of the aerial makes the greater contribution to the resultant field in the outermost minimum. Since the contributions from the two halves are in antiphase, the field strength will be increased if the inequality between the magnitudes of these contributions is increased. This is done by

Table 1

RELATIVE PHASES OF RADIATING CURRENTS AND CONTRIBUTIONS TO THE DISTANT FIELD

Tier	Without gap filling		With gap filling		
	Radiating current	Distant field at 17° to tower faces	Radiating current	Distant field at 17° to tower faces	
	deg	deg	deg	deg	
1 (top)	0	0	0	0	
2	0	0	0	0	
3	0	0	0	0	
4	0	0	0	0	
5	−72	0	−34	+38	Resultant = 0°
6	−89	0	−101	−12	
7	−106	0	−118	−12	
8 (bottom)	−123	0	−135	−12	

advancing the phase of the radiating current in the fifth tier (the highest tier in the lower half of the aerial), thereby tilting the maximum of the v.r.p. of the lower half downwards.

The calculated effect of the phase changes is illustrated by the full-line curves in Fig. 6. It will be seen that the minima are less deep than those in the broken curves, apart from one region

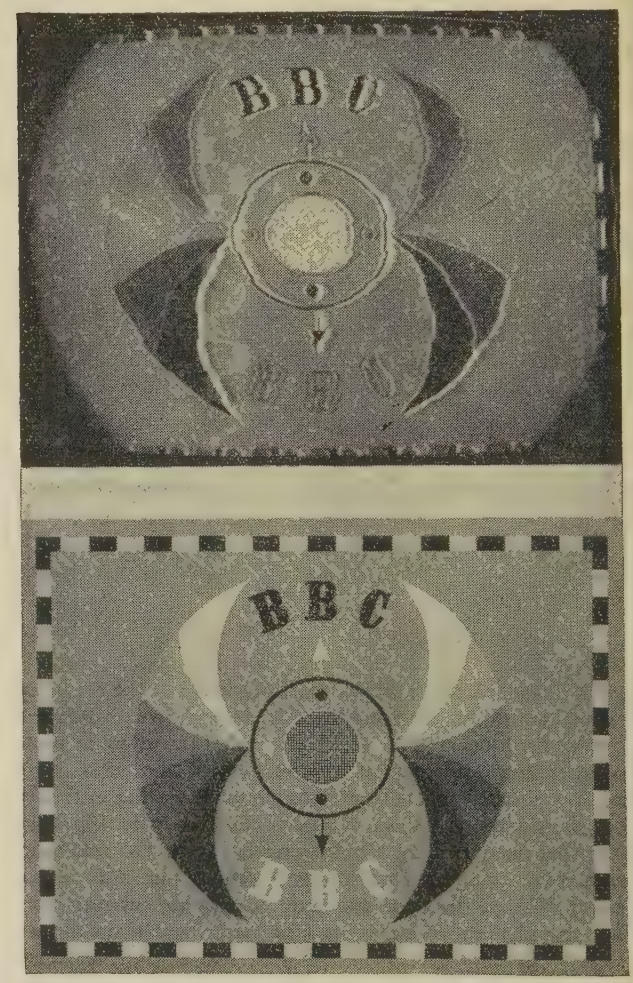


Fig. 9.—Distortion of test card.

(*a*) Received picture showing typical distortion.
(*b*) Transmitted test card.

close to the aerial. The improvement is illus-ed in a different manner in Fig. 8(*b*), a plot of field strength on the Argand diagram. The er half of Fig. 7 shows that the gap filling ild appreciably reduce the variation of field ngth at any one point over the band occupied he transmission.

) PERFORMANCE OF THE AERIAL WITH GAP FILLING

When the aerial system was first brought into ice, some measurements of field strength and rvations of picture quality were made in the ediate vicinity. The results indicated that the al was operating substantially as planned, but it ow realized that insufficient measurements were le to reveal the worst receiving conditions which ted. Complaints were received from viewers he immediate vicinity and it became apparent the gap filling was inadequate. Reception ditions at distances up to 1·5 miles were efore investigated by making measurements of strength simultaneously with observations of received picture. Severe distortion was found ight small areas (each about 400 yd long and yd wide) about one mile from the transmitting al in the outermost minimum of the v.r.p. ese areas correspond approximately to the −15 −20 dB contours in Fig. 10(*a*).] The principal ortion was in the tone rendering, which was so -linear that some pictures resembled photo-hic negatives. There was often difficulty in chronization. Fig. 9 shows the distortion of a card in a typical case.

was established that much of the distortion e from differences between the modulation acteristics of the two transmitters. These rences were very small, since the transmitters been built to a strict specification, framed with ration in parallel very much in mind. It would, ever, have been impracticable to frame the ification sufficiently strictly to provide for ons in which the two transmitters would con-ute to the received signal in antiphase, since small departures from perfection, if markedly rent as between one transmitter and another, ld be magnified many times. For example, if transmitter had only a slightly non-linear acteristic, the resultant would be much more -linear. On the other hand, if the ratio of the litude of the picture signal to that of the syn-nizing pulses were to differ slightly from one smitter to another, the resultant could exhibit ery low picture amplitude, or even a negative ure. Alternatively, the synchronizing pulses in resultant signal might be virtually non-existent. qual phase modulation, too slight to have any ct on the distant field where the contributions n the transmitters are nominally in phase, could lt in severe non-linear distortion when the con-utions are in antiphase and approximately equal agnitude. It was found that this last effect was most difficult to overcome, since it was pro-ndly influenced by very small changes in the ng of the r.f. input circuits of the modulated lifiers.

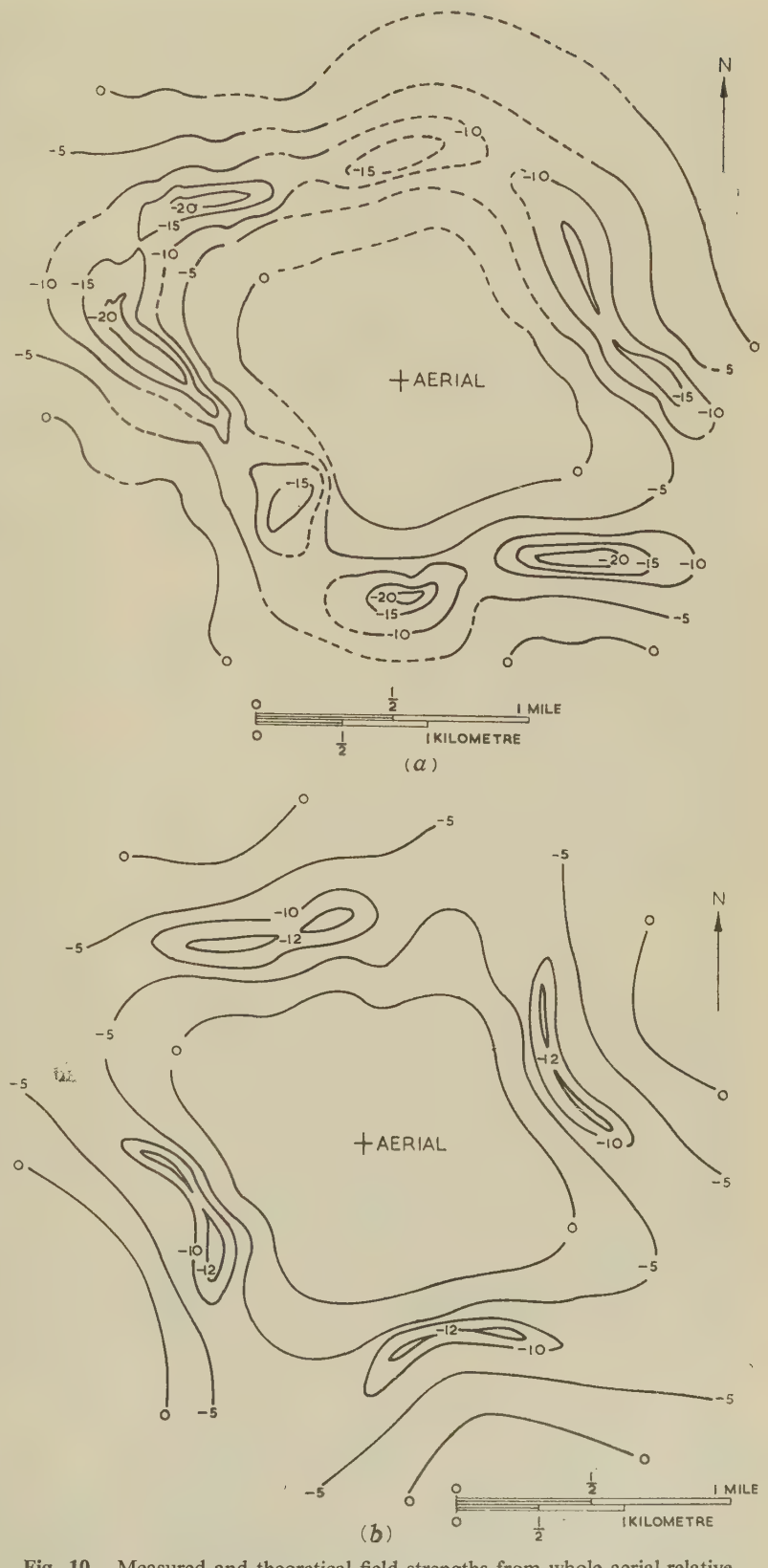


Fig. 10.—Measured and theoretical field strengths from whole aerial relative to that from single transmitter and lower half of aerial.

Numbers on contours are decibels.

(*a*) Measured.

—— Measured. - - - - Estimated (in inaccessible areas).

(*b*) Theoretical: no reflection at the ground.

One method used to investigate the effects described above was to simulate the receiving conditions at the transmitting station by taking signals from the outputs of the transmitters, combining them in antiphase with a known amplitude difference and examining the picture produced by the resultant signal. It was found that a difference in amplitude of at least 3 dB was required to give acceptable picture quality. It was possible to reduce this figure by careful adjustment of the transmitters, but to maintain this condition from day to day would have imposed an excessive burden on the operating staff.

In order to determine how nearly the contributions from the two halves of the aerial were cancelling one another, the field strength was measured with the whole aerial in use and compared with that obtained with the lower half only in use. The smaller the ratio of the field strength of the whole aerial to that of half the aerial, the greater the degree of cancellation. When the whole aerial was used, it was energized by one transmitter only (this facility was available as an emergency condition[6]), since for investigating the performance of the aerial it was desirable to eliminate any effect due to differences between the transmitters. Observations of picture quality were also made in this condition; the result corresponded to that which would be obtained if the transmitters could be made identical in performance.

Figs. 10(*a*) and (*b*), respectively, show measured and theoretical ratios of the field from the whole aerial to that from the lower half at a height of 15 ft above ground level. In Fig. 10(b) an approximate correction has been made for the ground contours by moving the field-strength contours outwards in proportion to the difference in height between the ground at each point and the centre of the aerial system.

A comparison between Figs. 10(*a*) and (*b*) shows quite good qualitative agreement. The subdivision of each of the four areas of low field strength is more pronounced in Fig. 10(*a*), but otherwise the positions of the minima are as predicted by the theory. Nevertheless the experimental results indicate that the contributions from the halves of the aerial can cancel one another to a greater degree than had been predicted. For example, the lowest ratio shown in Fig. 10(*b*) (theoretical), −12 dB, corresponds to contributions differing in amplitude by 2·5 dB, assuming them to be in antiphase. In the experimental contour map [Fig. 10(*a*)] the minimum ratio is about −24 dB, corresponding to a difference of only 0·5 dB. By varying the power in one half of the aerial it was established that the field from the lower half was the greater.

The picture quality, with only one transmitter in use, was generally satisfactory, but with the whole aerial in use, some distortion of the type associated with non-uniform frequency response could be seen in small regions where cancellation was almost complete. Reception was generally satisfactory except in the neighbourhood of the outermost minimum.

The conclusions may be summarized as follows. Even had the minima been filled in to the degree intended, it would have been impracticable to make the transmitters sufficiently similar in their characteristics to avoid distortion in regions where the contributions from the transmitters were in antiphase. But in fact, the outermost minimum was not filled in to the degree expected.

Since there was no reason to suspect that the aerial was behaving other than as planned, it appeared possible that the inadequacy of the gap filling might arise from propagation effects, such as ground reflection and diffraction over buildings. As stated in Section 4, reflection at the ground was neglected in the theoretical analysis since it was not thought possible to predict it in a built-up area. Calculations for a limited range of distances have, however, shown that if a reflection coefficient of −1 had been assumed, the agreement between theoretical and experimental results would have been much closer.

Since the halves of the transmitting aerial are at differe heights, their respective contributions will have different diffra tion angles at an obstacle, that from the lower half suffering t greater attenuation. This could be particularly significant regards the measurements of field strength, which were ma with a relatively low receiving aerial.

In order to investigate these propagation effects fully it wou be necessary to make extended observations with a high receiving aerial, but this was not practicable. At a few sites t receiving aerial was raised from 15 to 30 ft; the variations field strength indicated one or both effects to be significant, it was impossible to determine their relative importance.

(9) MODIFICATIONS

In the light of the results of observations referred to abov appeared impracticable to provide enough gap filling to overco the effect of differences between the modulation characteris of the transmitters without an appreciable reduction of fi strength over the greater part of the service area. It was theref decided to combine the outputs of the transmitters in a hy network, usually referred to in this application as a diple and then to divide the combined output between the halve the aerial. The effect of differences between the modul characteristics of the transmitters would then be the same a receiving points as it had been originally in the more di parts of the service area.

This arrangement entailed some loss of reliability, sinc diplexer was common to both halves of the transmitting insta tion, so that its failure would interrupt transmission. T consideration was not, however, considered important, si passive networks fail rarely. It is true that the repair of a work designed for high power may take a long time, bu is not necessary to take this into account provided that service can be restored in the meantime by the operation switches.

Although it was clear that the use of a diplexer would rem the cause of the great majority of complaints, it appea desirable at the same time to increase the field strength in sharp minima shown in Fig. 10(*a*), where some distortion of transmitted spectrum was apparent even when a single tra mitter was used. This was now easy, since for the first tim had become practicable to feed unequal powers to the halves of the aerial. It has already been stated that in th minima the contribution from the lower half of the aerial predominant, and it was therefore decided to feed the gre part of the total power to the lower half in order to accentu this predominance.

Consideration of the measured depth and sharpness of outermost minimum showed that, to reduce the variation field strength over the video band to less than 3 dB, it would necessary for the power in the lower half of the aerial to ex that in the upper half by 2 dB. The effect on minima close the base of the aerial was calculated to be beneficial; in any c observations had shown that these were already adequa filled, apart from one less than 300 ft from the tower wh there are no houses.

The unequal power division between the halves of the ae results in a decrease of 0·1 dB in the gain, averaged over horizontal directions. Since the h.r.p.'s of the halves di the radiation pattern of the whole aerial undergoes a s change, becoming slightly more uniform. Taking height- into account, the distant field is decreased by 0·3 dB for bear of 45° to the tower faces, i.e. for the directions of the max of the h.r.p. For bearings normal to the tower faces the negligible change in the distant field.

ig. 11.—Measured field strength from modified condition relative to field strength from single transmitter and lower half of aerial.

Numbers on contours are decibels.

——— Measured.
– – – – Estimated (in inaccessible areas).

(10) PERFORMANCE AFTER MODIFICATION

'hen the modifications to combine the outputs of the trans- ers and to divide the power in the required ratio were com- ed, the ratio of the power radiated by the upper half to that ated by the lower half was measured. The measured ratio, dB, was considered to be sufficiently close to the specified ie of 2 dB.

. survey of field strengths near the aerial gave results which plotted in contour form in Fig. 11 on the same basis as 10(*a*). It may be seen by comparing the two figures that minimum field strength has been increased by about 10 dB now approximates to the condition originally intended, wn in Fig. 10(*b*).

bservations of the quality of reception with an omnidirec- al aerial 14ft above ground level showed this to be satis- ory almost everywhere. The only remaining defect was the sence of weak delayed images in some small areas of relatively field strength. It was, however, considered that the use of mple directional aerial mounted somewhat higher would, in y case, give adequate reception.

(11) CONCLUSIONS

Experience at Crystal Palace has confirmed that, when a trans- ting aerial of high gain is erected in a populous area, great e is needed to ensure satisfactory reception in the imme- te vicinity. The difficulty of doing so is increased if the iation patterns of the separate tiers of the aerial are not all same, but this difficulty can be overcome and a satisfactory lt obtained. Some sacrifice in the distant field is entailed, but this sacrifice can be restricted to a fraction of a decibel. In computing the field strength theoretically it appears desirable to take ground reflection into account, ignoring the effect of buildings.

The use of entirely separate transmitters connected to different parts of the aerial is often the best and simplest method of ensuring transmissions uninterrupted by faults. Such a system should not, however, be used if there are receivers situated in the minima of the v.r.p. of the whole aerial, where contributions from the transmitters are received in antiphase. In this situation it is better to combine the outputs of the transmitters in a hybrid unit before redistributing it to parts of the aerial.

The problem of ensuring satisfactory reception in the immediate neighbourhood of a high-gain aerial will probably become more acute when Bands IV and V are used for television, since it will be necessary to use transmitting aerials of much higher gain. The methods used for gap filling at Crystal Palace will usually be inadequate in Bands IV and V, and more elaborate procedures such as have been employed in the U.S.A. will be necessary to give satisfactory local reception without undue sacrifice in the field strength at great distances. Fortunately, some of the practical problems will be eased in Bands IV and V. The effect of finite distance will be negligible and the fact that the transmitted band of frequencies is much narrower in relation to the carrier frequency can be used to simplify the design of the distribution feeder system.

(12) ACKNOWLEDGMENTS

A number of the authors' colleagues took part in the work described. In particular, Messrs. R. V. Harvey and D. W. Osborne carried out much of the early theoretical work and devised the method of presentation used in Fig. 6. Computation was performed by Miss G. M. Oulton and Mr. K. W. T. Hughes. Field-strength measurements were undertaken by Mr. D. W. Taplin. The engineering design of the feeder system at Crystal Palace, including implementation of the gap-filling measures, was carried out by the Planning and Installation Department of the British Broadcasting Corporation, which also made detailed measurements of the linearity and phase modulation of the transmitters. The authors are indebted to the Director of Engineering of the B.B.C. for permission to publish the paper.

(13) REFERENCES

(1) KEAR, F. G., and PRESTON, J. G.: 'Control of Vertical Radiation Patterns of T.V. Transmitting Antennas', *Proceedings of the Institute of Radio Engineers*, 1954, **42**, p. 402.

(2) MCLEAN, F. C., THOMAS, A. N., and ROWDEN, R. A.: 'The Crystal Palace Television Transmitting Station', *Proceedings I.E.E.*, Paper No. 2069 R, March, 1956 (**103** B, p. 633).

(3) WOODWARD, P. M.: 'A Method of Calculating the Field over a Plane Aperture Required to Produce a Given

Polar Diagram', *Journal I.E.E.*, 1946, **93**, Part IIIA, p. 1554.

(4) BROWN, G. H.: 'Pattern Synthesis—Simplified Methods of Array Design to Obtain a Desired Directive Pattern', *RCA Review* 1959, **20**, p. 398.

(5) NEWTON, I.: ' "Close-in" Coverage with High-Gain V.H.F. Antennas', *Broadcast News*, September–October, 1952, p. 44.

(6) WHARTON, W., and PLATTS, G. C.: 'The Crystal Palace Band I Television Transmitting Aerial', B.B.C. Engineering Division Monograph No. 23, February, 1959.

(7) MONTEATH, G. D., WHYTHE, D. J., and HUGHES, K. W. T 'A Method of Amplitude and Phase Measurement in th V.H.F.–U.H.F. Band', *Proceedings I.E.E.*, Paper N 3155 E, March, 1960 (**107** B, p. 150).

(8) CARTER, P. S.: 'Antenna Arrays Around Cylinders', *Pro ceedings of the Institute of Radio Engineers*, 1943, 3 p. 671.

.391.82 : 621.396.65

The Institution of Electrical Engineers
Paper No. 3326 E
Jan. 1961

RADIO-FREQUENCY INTERFERENCE IN MULTI-CHANNEL TELEPHONY F.M. RADIO SYSTEMS

By R. HAMER, Ph.D., B.Sc., Associate Member.

(*The paper was first received 4th February, and in final form 27th June*, 1960.)

SUMMARY

An investigation is described of the effect of extraneous radio-quency interference in frequency-modulation multi-channel radio-phony transmission systems. The noise power ratio in the output phony baseband due to an r.f. interfering signal is derived oretically, and comparison is made with measured values. When practical limitations in the method of measurement are taken o account, it is concluded that the theoretical results are valid. Results of general application in the design of radio systems are rived from the theory, and the analogy, under certain circumstances, h interference due to long-delay echo-signals is discussed.

LIST OF PRINCIPAL SYMBOLS

$a = \frac{1}{2}(f_s - f_r)$.
$b = \frac{1}{2}(f_s + f_r)$.
δf = Infinitesimally small frequency band; also measuring bandwidth.
$\delta x = \delta f / f_n$.
ΔF = Total r.m.s. frequency deviation of f.m. signal (ΔF_1, wanted signal; ΔF_2, interfering signal).
Δf_i = R.M.S. frequency deviation imparted to wanted carrier by interfering signal, corresponding to interference power in a small band at f_r in the baseband.
Δf_s = R.M.S. deviation of wanted carrier corresponding to the signal power in a small band at f_r in the baseband.
f = Frequency relative to an arbitrary zero.
f_1 = Minimum modulation frequency of wanted signal.
f_2 = Minimum modulation frequency of interfering signal.
f_m = Maximum modulation frequency of interfering signal.
f_n = Maximum modulation frequency of wanted signal.
f_r = Instantaneous frequency separation of wanted and interfering signals; also, mean baseband frequency of small band under consideration.
f_s = Mean frequency separation of wanted and interfering signals.
I = General spectrum integral in f.
I_x = General spectrum integral in x.
$I'_x = I_a + I_b$ = particular spectrum integral in x.
m = R.M.S. modulation index $= \Delta F / f_n$ (m_1, wanted signal; m_2 interfering signal).
$m'_2 = \Delta F_2 / f_n$.
$\mathscr{P}$ = Interference probability.
ϕ = Relative phase of resultant vector on addition of interfering signal.
u = Ratio of amplitudes of wanted and interfering signals $(= V_s/V_i)$.
V_i = Amplitude of interfering signal.
V_r = Amplitude of resultant of wanted and interfering signals.
V_s = Amplitude of wanted signal.
P = Total power of wanted baseband signal.
P_D = Normalized residual carrier power.
P_i = Mean interference power in a small band at f_r in the baseband.
P'_i = Instantaneous interference power at f_r in the baseband.
$P(f)$ = F.M. mean noise-power spectrum with unit carrier power. [$P(x)$ in terms of the variable x.]
P_s = Mean wanted signal power in a small band at f_r in the baseband.
$x = f / f_n$.
$x_1 = f_1 / f_n$.
$x_2 = f_2 / f_m$.
$x'_2 = f_2 / f_n$.
$x_a = a / f_n = \frac{1}{2}(x_s - x_r)$.
$x_b = b / f_n = \frac{1}{2}(x_s + x_r)$.
$x_m = f_m / f_n$.
$x_r = f_r / f_n$.
$x_s = f_s / f_n$.

(1) INTRODUCTION

One of the sources of noise to be considered in frequency-division multiplex radio systems using frequency modulation (f.m.) is that due to extraneous radio-frequency (r.f.) interference. This problem was first investigated in 1955, when some brief results were described.[1] A more thorough investigation was later carried out independently by Medhurst *et al.*[2]

In the earlier investigation three types of interfering signal were considered:

(*a*) An unmodulated signal.
(*b*) A similar f.m. signal.
(*c*) An f.m. television signal.

The present investigation is confined to types (*a*) and (*b*), and forms an extension of the previous work.* The scope and treatment differ from those of Reference 2, and graphical results are provided in generalized form.

As in the previous investigations, the multi-channel signal has been simulated by a random-noise signal of uniform power density. The results are given in terms of the *noise power ratio* (n.p.r.), from which the test-tone/noise ratio in the telephone channels of a practical system can readily be deduced.[3] Although the representation of a multi-channel signal by a random-noise signal is permissible only in high-capacity systems, the calculation or measurement of the n.p.r. in low-capacity systems is of value. For, even when a few channels are involved, the signal/noise ratio in any channel can still be related empirically or statistically to the n.p.r.

(2) THEORY

(2.1) General

In order to isolate the problem under investigation it is convenient to consider an ideal f.m. radio-communication system, free from all forms of distortion and basic noise. An r.f. carrier wave is assumed to be frequency modulated by a

Written contributions on papers published without being read at meetings are vited for consideration with a view to publication.
The paper is based on a thesis approved for the Ph.D. degree of the University of ndon.
Dr. Hamer is in the Research Branch of the Post Office Engineering Department.

* The author is indebted to Mr. Medhurst for drawing his attention to some errors in the earlier analysis.

uniform-spectrum noise signal occupying a finite baseband. After transmission, the noise signal is recovered from an ideal f.m. receiver. The n.p.r. at any frequency in the baseband is therefore initially infinite. If an extraneous signal is now introduced into the receiver, spurious noise will be added to the demodulated signal, and a finite n.p.r. will result. The mechanism may be understood by first considering the effect of an unmodulated interfering signal within the spectrum of the wanted signal, and of much smaller amplitude than the wanted signal [Fig. 1(*a*)]. Whenever the wanted carrier momentarily

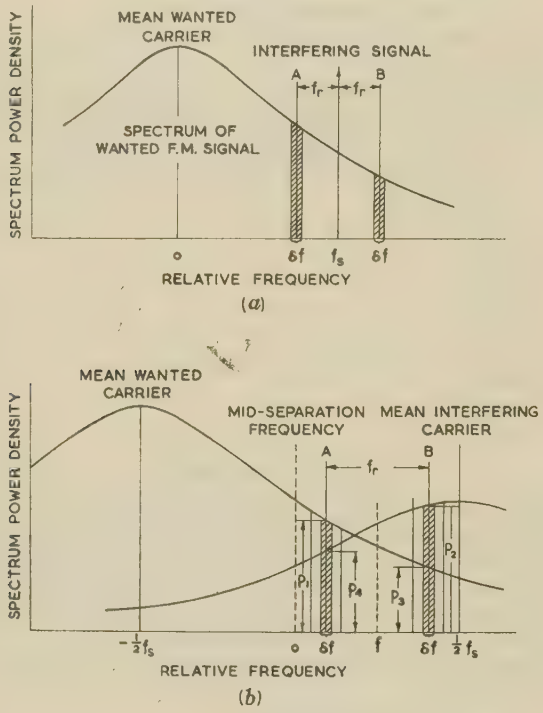


Fig. 1.—Spectrum diagrams.
(*a*) Unmodulated interfering signal.
(*b*) Noise-modulated interfering signal.

occupies one of the small bands, δf, separated by a frequency f_r from the interfering signal, the wanted carrier is phase and amplitude modulated at the frequency f_r. The amplitude modulation is suppressed by the limiters in the ideal demodulator, but the phase modulation appears as frequency modulation at the discriminator. Under these conditions a spurious signal of frequency f_r is therefore generated in the output of the demodulator. Since the wanted carrier is modulated by a random-noise signal, it is reasonable to suppose that the bands, δf, are occupied in random manner; consequently the interference power in a small band near f_r in the baseband is expected to vary in random manner. In an actual multi-channel system, unintelligible interference, or noise, is in fact experienced. However, in certain circumstances, as pointed out by Medhurst *et al.*,[2] and discussed further in Appendix 8.1, some intelligible* crosstalk can, in theory, occur.

It has been shown[4, 5] that, when quasi-stationary conditions[6] apply, the carrier resides in a given frequency band for a proportion of the time (or with a probability) that is equal to the mean spectrum power in the band, normalized for unit carrier power. Thus, referring to Fig. 1(*a*), the carrier occupies each of the very small bands, δf, with a probability equal to the product of δf and the normalized mean power density of the spectrum at the frequency corresponding to each small band. While the carrier occupies one of these infinitesimal bands, the spurious frequency modulation created by the interfering signal is readily calculated. The mean interference power density at frequency f_r in the baseband is therefore the product of the calculated interference power and the sum of the mean power densities at A and B in the spectrum. This method may be extended, as described in Section 2.4, to include the effect of interference from a noise-modulated f.m. signal. However, owing to the initial assumption, the method is only applicable strictly to systems with large modulation indices.

The method may be extended to include f.m. systems with any value of modulation index if the following hypothesis is made: the probability that the instantaneous frequency* of the f.m. signal occupies a given frequency band is equal to the normalized mean power of the spectrum in that band. This hypothesis, which has not been encountered in the literature, is adopted here, subject to confirmation by the results of measurements.

From the outline of the method so far described, it may be concluded that the interference noise power density in the baseband is dependent on:

(*a*) Relative amplitude of interfering signal.
(*b*) Modulating-signal frequency band.
(*c*) Particular baseband frequency considered.
(*d*) Mean frequency separation between wanted and interfering signals.
(*e*) Mean noise power spectrum of the wanted signal.
(*f*) Type of interfering signal, and, if it is a similar noise-modulated f.m. signal, mean noise power spectrum of interfering signal.

Since only small-amplitude interfering signals are of importance in practice, it is convenient to restrict the analysis to this case. The effect of an interfering signal comparable in amplitude with the wanted signal is, however, discussed in Appendix 8.2.

The f.m. noise power spectrum always extends theoretically over an infinite frequency band. The ideal receiver is therefore assumed to have infinite bandwidth. It also follows that in an ideal receiver the carrier frequency must be very high compared with both the maximum modulation frequency and the r.m.s. frequency deviation. These conditions normally apply in practical f.m. systems, since they are essential requirements for low distortion. No restriction is placed on the carrier frequency other than this, and, in the analyses, all frequencies are specified relative to an arbitrary reference. Since the f.m. noise power spectrum is symmetrical about the mean carrier frequency, only positive values of frequency separation need be considered.

The importance of the f.m. mean noise power spectrum in the present analysis is such that it is treated separately in the Subsection following.

(2.2) The F.M. Noise Power Spectrum

Several analyses of the f.m. noise power spectrum[4, 7, 8] have been carried out, and measurements have also been described.[5]

It is convenient to summarize here the limiting approximate formulae for the f.m. noise power spectrum for very small and for large modulation indices. Dimensionless variables are used as in Reference 5, and the results are normalized for unit carrier power. Since the modulating signal is a random-noise voltage, the modulation index cannot usefully be defined as the ratio of the peak deviation to the highest modulation frequency. Instead an *r.m.s. modulation index*, defined as the ratio of the r.m.s. deviation to the highest modulation frequency, is introduced

* i.e. coherent interference, such that recognizable speech, etc., from another channel or channels may be heard.

* Instantaneous angular frequency is defined as rate of change of phase (see, for example, Reference 6).

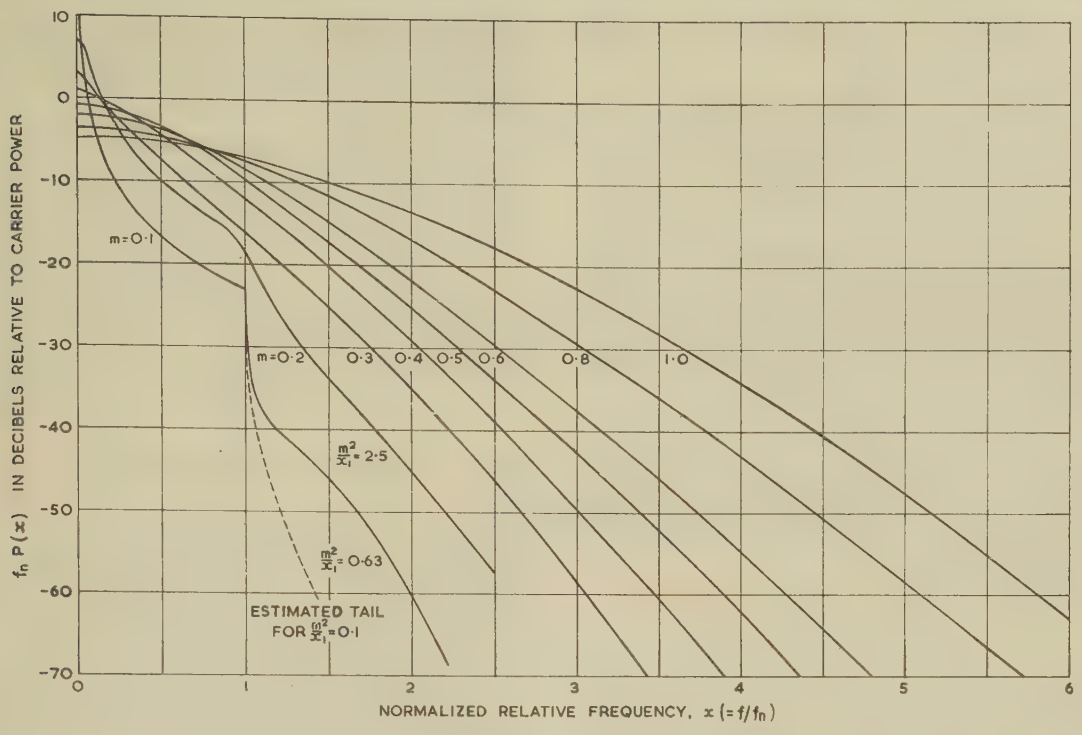


Fig. 2.—Normalized f.m. noise power spectra.

only does this avoid the use of an arbitrary peak factor, it also describes the modulation index in terms of measurable ameters.

When the r.m.s. modulation index is very small (less than ut 0·2) the shape of the spectrum is dependent on the lowest dulation frequency. It has been shown[9] that, if the lowest uency is zero, an unbounded continuous spectrum results:

$$f_n P(x) \simeq \frac{2m^2}{\pi^2 m^4 + 4x^2} \quad . \quad . \quad . \quad . \quad . \quad (1)$$

lowever, this cannot arise in practice, since the lowest dulation frequency cannot be zero. When the lowest fre- ncy is not zero, a bounded continuous spectrum results, ether with a residual carrier at the mean carrier frequency. normalized residual carrier power, P_D, given by Middleton,*

$$P_D = \exp(-m^2/x_1) \quad . \quad . \quad . \quad . \quad . \quad (2)$$

he continuous spectrum is then[10]

$$f_n P(x) = \frac{m^2}{2x^2(1-x_1)} \ldots [x_1 \leqslant |x| \leqslant 1] \quad . \quad . \quad (3)$$

he residual carrier corresponds to the carrier component of spectrum when a single modulating tone is used. It may be umed that the instantaneous frequency remains precisely at mean carrier frequency with a probability given by eqn. (2). s. (1)–(3) are approximations for the small modulation-index se power spectrum; as measurement confirms,[5] the actual trum falls between these limiting forms. It is suggested in erence 5 that the value of m^2/x_1 is a useful criterion; the ller this ratio, the more closely does the continuous spectrum roach the form of eqn. (3).

When the r.m.s. modulation index is large (greater than about 1·5) the mean power spectrum down to normalized power densities of at least −60 dB is of Gaussian form:[5]

$$f_n P(x) \simeq \frac{1}{\sqrt{(2\pi)m}} \exp(-x^2/2m^2) \quad . \quad . \quad . \quad (4)$$

The residual carrier is now negligible, the spectrum being continuous and unbounded, and substantially independent of the value of x_1.

For intermediate values of m, noise power spectra based on measurement are believed to be the most reliable, and in the present work the measurements described in Reference 5 are used. Normalized spectrum curves obtained from these measurements are shown in Fig. 2, for eight values of m between 0·1 and 1·0.

(2.3) Unmodulated Interfering Signal

(2.3.1) General.

Referring to Fig. 1(*a*), consider an interfering signal within the spectrum of a noise-modulated f.m. signal, separated by a frequency f_s from the mean carrier. Let the instantaneous frequency of the f.m. signal occupy one of the infinitesimal bands, δf, separated by a frequency f_r from the interfering signal. The two r.f. signals are applied additively to the ideal f.m. receiver, the instantaneous signal vectors being as depicted in Fig. 3. If the amplitude of the interfering vector, V_i, is small compared with that of the signal vector, V_s, the spurious angular modulation imparted to the resultant carrier vector is

$$\frac{d\phi}{dt} \simeq \frac{2\pi f_r}{u} \cos 2\pi f_r t \quad . \quad . \quad . \quad . \quad . \quad (5)$$

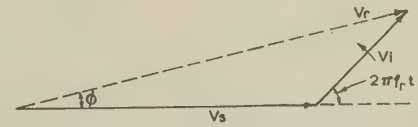


Fig. 3.—Vector diagram.

* An error in the exponent was pointed out by Stewart.

where u is the ratio V_s/V_i. The amplitude modulation impressed upon the resultant has been neglected, since this is removed in the amplitude limiter in the receiver.

Eqn. (5) shows that the frequency of the spurious frequency modulation is f_r, as would be expected, and that the r.m.s. value of the spurious frequency deviation, Δf_i, is

$$\Delta f_i \simeq f_r/\sqrt{2}u \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

The interference power after demodulation, in a small band δf, at a mean frequency f_r in the baseband, may now be obtained as follows. Let the total wanted mean noise power in the baseband be P, and in the small band δf, P_s; then the r.m.s. deviation, Δf_s, of the wanted carrier, required to produce the power P_s, is given by

$$\Delta f_s^2 = \Delta F^2 P_s/P \, . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

where ΔF is the r.m.s. deviation produced by the total noise signal. The ratio of the instantaneous interference power, P_i', to the mean wanted noise power, P_s, in the small band δf in the baseband is therefore

$$\frac{P_i'}{P_s} = \frac{\Delta f_i^2}{\Delta f_s^2} \simeq \frac{f_r^2}{2u^2\Delta F^2} \frac{P}{P_s}$$

from eqns. (6) and (7). If we now specify a noise signal spectrum extending from a frequency f_1 to a frequency f_n in the baseband, we may write

$$\frac{P_i'}{P_s} = \frac{f_r^2(f_n - f_1)}{2u^2\Delta F^2 \delta f} \, . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

Eqn. (8) gives the ratio of instantaneous interference to mean signal power in a small band, δf, at a frequency f_r in the baseband, during the time that the interfering signal occupies one of the two bands, δf, shown in Fig. 1(*a*).

In order to obtain the mean interference/signal power ratio (i.e. the reciprocal of the n.p.r.) it is necessary to multiply P_i'/P_s by the probability the f.m. carrier has of occupying the two bands, δf. From the initial hypothesis these probabilities are

$$\mathscr{P}_1 = P(f_s - f_r)\delta f$$

for the band at A in the diagram, and

$$\mathscr{P}_2 = P(f_s + f_r)\delta f$$

for the band at B, where $P(f)$ is the f.m. mean noise power spectrum, normalized for unit carrier power. The required result is therefore

$$\frac{P_i}{P_s} = \frac{P_i'}{P_s}(\mathscr{P}_1 + \mathscr{P}_2)$$

where P_i is the mean interference power in the band δf at f_r in the baseband. Thus, using eqn. (8), we have for the n.p.r. ($= P_s/P_i$) at f_r in the baseband

$$(\text{n.p.r.})_r \simeq \frac{2u^2\Delta F^2}{f_r^2(f_n - f_1)} \Big/ [P(f_s - f_r) + P(f_s + f_r)] \quad . \quad (9)$$

Eqn. (9) may be expressed in terms of the reduced variables as follows:

$$(\text{n.p.r.})_r \simeq \frac{2u^2m^2}{x_r^2(1 - x_1)} \Big/ [f_nP(x_s - x_r) + f_nP(x_s + x_r)] \quad . \quad (10)$$

It may be seen from eqns. (9) and (10) that, subject to the assumption that u is large, the n.p.r. is directly proportional to u^2. The dependence of the n.p.r. on the remaining variables cannot be predicted without a knowledge of the noise pow spectrum. The result includes any intelligible interference pow since all possible conditions for interference to arise have be included.

(2.3.2) Calculation for Very Small Modulation Index.

The asymptotic form of the f.m. noise power spectrum wh the modulation index is very small is given in eqns. (1)–(3). practice, the value of m^2/x_1 is usually small, so that eqns. and (3) provide the better approximation to the spectrum.

The interference n.p.r., at a frequency f_r in the baseband, d to the continuous portion of the spectrum, is readily obtain from eqns. (3) and (10):

$$(\text{n.p.r.})_r = \frac{4u^2}{x_r^2} \Big/ \left[\frac{1}{(x_s - x_r)^2} + \frac{1}{(x_s + x_r)^2}\right] \quad .$$

where $$x_1 \leqslant |x_s - x_r| \leqslant 1$$

and $$x_1 \leqslant (x_s + x_r) \leqslant 1$$

only positive values of x_s being considered.

When $x_s = x_r$, i.e. $f_s = f_r$, in addition to a possi contribution from the second spectral term in eqn. (11) $x_r \leqslant \frac{1}{2}$), interference is imparted to the residual carrier t frequency f_r. Regarded as a spectral component, the resi carrier takes the form of a delta function, and therefore gi rise to an infinite interference noise power density at f_r in baseband. The n.p.r. for this frequency separation is there determined by the measuring bandwidth. Thus, regarding (2) as expressing the probability that the f.m. signal reside the mean carrier frequency, the n.p.r. when $f_s = f_r$ may derived from eqn. (8) (transformed to reduced variables) a eqn. (11). The result is

$$(\text{n.p.r.})_r \simeq \frac{4u^2}{x_r^2} \Big/ \left[\frac{2(1 - x_1)}{m^2\delta x}\left(\exp - \frac{m^2}{x_1}\right) + \frac{1}{4x_r^2}\right] \quad .$$

where δx may now be regarded as the normalized measuri bandwidth, it being noted that $x_s = x_r$, and that the seco term in the denominator becomes zero when $x_r > \frac{1}{2}$.

Excluding the effect of the residual carrier, it may be seen inspection of eqn. (11) that the n.p.r. is lowest when x_s has its smallest value; i.e.

when $$x_s = x_r + x_1$$

When $x_s > 1$, the n.p.r. becomes theoretically infinite; practice it is expected to become very large, being determi by the tails of the f.m. noise spectrum. It may also be not from eqn. (11) that the n.p.r. is independent of the modulati index, provided that it is very small, as assumed.

(2.3.3) Calculation for Large Modulation Index.

When the modulation index is large, the mean noise pow spectrum approaches the Gaussian form given in eqn. Inserting this result in eqn. (10), the following is obtained the n.p.r. at a frequency f_r in the baseband:

$$(\text{n.p.r.})_r \simeq \frac{2u^2m^2/[x_r^2(1 - x_1)]}{\frac{1}{\sqrt{(2\pi)}m}\left[\exp - \frac{(x_s - x_r)^2}{2m^2} + \exp - \frac{(x_s + x_r)^2}{2m^2}\right]} \quad . \quad (1$$

This equation may be rearranged as follows:

$$(\text{n.p.r.})_r \simeq \sqrt{(2\pi)} \frac{u^2m^3}{x_r^2(1 - x_1)} \operatorname{sech} \frac{x_sx_r}{m^2} \exp \frac{(x_s^2 + x_r^2)}{2m^2} \quad . \quad (1$$

Unlike the small-deviation approximation, the spectrum and ce the n.p.r. as a function of x_s are continuous, and it may seen from eqns. (13) that the n.p.r. varies directly as x_s. The .r. is approximately inversely proportional to x_r^2, since the ct of x_r in the exponential terms is insignificant. The n.p.r. nerefore always lowest at the highest baseband frequency. It y also be concluded that, for small values of x_s, the n.p.r. is roximately directly proportional to m^3, and hence to the e of the r.m.s. frequency deviation.

.4) **Computation from Noise Spectra.**

When an analytic expression for the f.m. noise power spectrum not be derived, use may be made of eqn. (10) to compute the .r. function from measured spectra. Alternatively, spectrum a computed from series approximations or other methods y be used.

(2.4) Noise-Modulated Interfering Signal

.1) **General.**

The theoretical analysis may be extended in the manner lowing to include an interfering signal that is frequency dulated by a random-noise signal.

A diagram depicting the overlapping spectra of the wanted and erfering signals is shown in Fig. 1(*b*). It is now convenient choose the arbitrary zero frequency midway between the an carrier frequencies. When two infinitesimal bands, δf, considered (at A and B in the diagram), separated by a quency f_r, account must be taken of interference falling in a all band $f_r \pm \delta f$ in the baseband. When the f.m. signals cupy the bands at A and B, the ratio of the instantaneous erference power to mean signal power must therefore be nsidered in a band $2\delta f$ at f_r in the baseband. The analysis ding to eqn. (8) is consequently modified by the substitution $2\delta f$ for δf, and we have

$$\frac{P_i'}{P_s} \simeq \frac{f_r^2(f_n - f_1)}{4u^2\Delta F_1^2 \delta f} \quad . \quad . \quad . \quad . \quad . \quad (14)$$

ere ΔF_1 is the r.m.s. frequency deviation of the wanted carrier. When the interfering signal is frequency modulated, there is infinite number of small bands δf, spaced f_r apart, to be nsidered. In the diagram, two such bands only are shown, th a mean frequency f. To include all interference conditions, e variable f must take all values between $-\infty$ and $+\infty$, and integration is therefore involved.

Consider any value of f, as shown in the diagram. The wanted rrier then occupies the band δf at A with a probability

$$\mathscr{P}_1 = P_1(f + a)\delta f \quad . \quad . \quad . \quad . \quad . \quad (15)$$

ere $P_1(f)$ is the normalized power spectrum of the wanted gnal. The value of $\mathscr{P}_1$ is directly proportional to the ordinate in the diagram. Similarly, the probability that the interfering rrier occupies a band δf at B is

$$\mathscr{P}_2 = P_2(f - a)\delta f \quad . \quad . \quad . \quad . \quad . \quad (16)$$

here $P_2(f)$ is the normalized noise power spectrum of the terfering signal. The value of $\mathscr{P}_2$ is thus proportional to the dinate p_2 in the diagram. For the appropriate interference nditions to be met, both small bands must be occupied multaneously, so the required probability is $\mathscr{P}_1\mathscr{P}_2$.

When both spectra are continuous and unbounded, and also other cases for some values of f, the inverse occupancy of the nds at A and B must be considered. The appropriate proba- lities are then proportional to the ordinates p_3 and p_4 in the agram, and are

$$\mathscr{P}_3 = P_1(f + b)\delta f \quad . \quad . \quad . \quad . \quad . \quad (17)$$

$$\mathscr{P}_4 = P_2(f - b)\delta f \quad . \quad . \quad . \quad . \quad . \quad (18)$$

The interference probability for the one pair of bands considered, at A and B, is therefore $(\mathscr{P}_1\mathscr{P}_2 + \mathscr{P}_3\mathscr{P}_4)$.

So far, only the small bands shown cross-hatched in the diagram have been considered. These are *full-probability* bands, in that, when they are occupied, the probability of interference falling in a band $2\delta f$ at f_r in the baseband is unity. However, when the wanted carrier is in the cross-hatched band at A, there are two *partial-probability* bands of width δf, adjacent on either side of the shaded band at B, to be considered. It is shown in Appendix 8.3 that, when each of these adjacent small bands is occupied, the interference probability is $\frac{1}{2}\mathscr{P}_1\mathscr{P}_2$. The interference probability for both adjacent bands at B is therefore $\mathscr{P}_1\mathscr{P}_2$, so the original probability $\mathscr{P}_1\mathscr{P}_2$ must be doubled. When the inverse band occupancy is considered, the adjacent bands at A must be taken into account, so the probability $\mathscr{P}_3\mathscr{P}_4$ must also be doubled. The total instantaneous interference probability, which is a function of f, is therefore

$$\mathscr{P}(f) = 2(\mathscr{P}_1\mathscr{P}_2 + \mathscr{P}_3\mathscr{P}_4)$$

and, from eqns. (15)–(18), this is

$$\mathscr{P}(f) = 2\delta f^2[P_1(f + a)P_2(f - a) + P_1(f + b)P_2(f - b)] \quad . \quad (19)$$

The required total mean interference probability, $\mathscr{P}$, is obtained by integration of the function $\mathscr{P}(f)$ over the whole of the frequency spectrum. Thus, as δf tends to zero, we may write

$$\mathscr{P} = \int_{-\infty}^{\infty} \mathscr{P}(f)\frac{1}{\delta f}df$$

i.e. using eqn. (19),

$$\mathscr{P} = 2\delta f \int_{-\infty}^{\infty} [P_1(f + a)P_2(f - a) + P_1(f + b)P_2(f - b)]df \quad . \quad (20)$$

The ratio of the mean interference/signal power in a small band $2\delta f$ at f_r in the baseband; i.e. in the limit the reciprocal of the n.p.r. at f_r is

$$\frac{P_i}{P_s} = \frac{P_i'}{P_s}\mathscr{P}$$

Hence, the n.p.r. at f_r in the baseband is, from eqns. (14) and (20),

$$(\text{n.p.r.})_r \simeq \frac{2u^2\Delta F_1^2}{f_r^2(f_n - f_1)I} \quad . \quad . \quad . \quad . \quad (21a)$$

where $I = \int_{-\infty}^{\infty} [P_1(f + a)P_2(f - a) + P_1(f + b)P_2(f - b)]df$

In terms of the reduced variables this result becomes

$$(\text{n.p.r.})_r \simeq \frac{2u^2 m_1^2}{x_r^2(1 - x_1)I_x} \quad . \quad . \quad . \quad . \quad (21b)$$

$$I_x = \int_{-\infty}^{\infty} [f_n P_1(x + x_a)f_n P_2(x - x_a) + f_n P_1(x + x_b)f_n P_2(x - x_b)]dx$$

It may be seen from eqns. (21) that, as with an unmodulated interfering signal, the n.p.r. is directly proportional to u^2, subject to the assumption that u is large. In this case also, the dependence of the n.p.r. on the remaining variables cannot be predicted without a knowledge of the noise power spectra. The result given in eqns. (21) includes the effect of intelligible interference.

(2.4.2) **Calculation for Very Small Modulation Indices.**

When the r.m.s. modulation indices of both the wanted and the interfering f.m. signals are very small, the normalized continuous noise power spectra may be represented by eqn. (3); thus, for the wanted signal,

$$f_n P_1(x) = \frac{m_1^2}{2x^2(1 - x_1)} \qquad [x_1 \leqslant |x| \leqslant 1] \quad . \quad . \quad (22)$$

For the interfering signal, it is convenient to use the following:

$$f_n P_2(x) = \frac{m_2'^2}{2x^2(x_m - x_2')} \qquad [x_2' \leqslant |x| \leqslant x_m] \quad . \quad (23)$$

The modulation index of the interfering signal, i.e. $\Delta F_2/f_m$, must be very small, but the parameter m_2' need not necessarily be so.

When these expressions are inserted in eqn. (21*b*) the following result is obtained for the n.p.r. at f_r in the baseband, due to the interaction of the continuous portions of the power spectra:

$$(\text{n.p.r.})_r \simeq \frac{8u^2(x_m - x_2')}{x_r^2 m_2' I_x'} \quad . \quad . \quad . \quad . \quad (24)$$

where $$I_x' = \int \left[\frac{1}{(x + x_a)^2(x - x_a)^2} + \frac{1}{(x + x_b)^2(x - x_b)^2} \right] dx$$

and the limits of integration have yet to be defined. The terms in the integral are similar, and if $I_x' = I_a + I_b$, then, on performing the integration and neglecting the constant of integration, we have

$$I_a = \frac{1}{4x_a^3} \left[\log \left| \frac{x + x_a}{x - x_a} \right| - \frac{x_a}{x + x_a} - \frac{x_a}{x - x_a} \right]$$

and I_b is the same, with x_b in place of x_a. From eqn. (21*b*) we may identify $(x + x_a)$ with the power spectrum of the wanted signal, and $(x - x_a)$ with that of the interfering signal. Noting the limits of these spectra in eqns. (22) and (23), the integration limits for I_a are seen to be

$$x_1 \leqslant |x + x_a| \leqslant 1$$

and

$$x_2' \leqslant |x - x_a| \leqslant x_m$$

and correspondingly for I_b, with x_b in place of x_a. Consideration of the geometry of the problem shows that the required integration is to be taken over the overlapping limits of x derived from the spectrum limits. The limits of x are

(i) $(-x_a - 1)$ to $(-x_a - x_1)$; $(x_1 - x_a)$ to $(1 - x_a)$
(ii) $(x_a - x_m)$ to $(x_a - x_2')$; $(x_a + x_2')$ to $(x_a + x_m)$

and similarly for I_b. Thus, the integral I_a is to be evaluated over the overlapping range of the limits (i) and (ii); and similarly for I_b.

The n.p.r. due to the continuous portions of the power spectra is therefore

$$(\text{n.p.r.})_r \simeq \frac{32u^2(x_m - x_2')}{x_r^2 m_2'^2} \Big/ \left[\frac{1}{x_a^3} \left(\log \left| \frac{x + x_a}{x - x_a} \right| - \frac{x_a}{x + x_a} - \frac{x_a}{x - x_a} \right) + \frac{1}{x_b^3} \left(\log \left| \frac{x + x_b}{x - x_b} \right| - \frac{x_b}{x + x_b} - \frac{x_b}{x - x_b} \right) \right] \quad . \quad (25)$$

with the terms in the denominator evaluated over the overlapping limits of x described. It may be noted that, when x_a or x_b is zero, the appropriate term in the denominator has the limiting value $2/3x^3$.

The n.p.r. due to the continuous spectra is degraded by further interference created by the interaction of the spectra with the residual carriers.

Considering first the effect of the interfering-signal residual carrier on the wanted signal, we may regard the residual carrier as an unmodulated interfering signal and apply the results of Section 2.3.2. Using eqn. (2), the power of the residual carrier relative to the full power of the interfering signal may be expressed

$$P_{D2} = \exp - \frac{m_2'^2}{x_2' x_m}$$

Thus, from eqn. (11), the n.p.r. from this cause is

$$(\text{n.p.r.})_r \simeq \frac{4u^2}{x_r^2} \exp \frac{m_2'^2}{x_2' x_m} \Big/ \left[\frac{1}{(x_s - x_r)^2} + \frac{1}{(x_s + x_r)^2} \right] \quad . \quad . \quad (26)$$

In considering the effect of the residual carrier of the wanted signal, it is more useful to regard this residual carrier as of full carrier power but residing at the mean carrier frequency for a proportion of the time given by eqn. (2). During this time, the n.p.r. due to the continuous spectrum of the interfering signal is given by eqn. (11), modified by the substitution of the interfering signal spectrum densities in the denominator. Thus, the n.p.r. from this cause is

$$(\text{n.p.r.})_r \simeq \frac{4u^2 m_1^2 (x_m - x_2')}{x_r^2 m_2'^2 (1 - x_1)} \exp \frac{m_1^2}{x_1} \Big/ \left[\frac{1}{(x_s - x_r)^2} + \frac{1}{(x_s + x_r)^2} \right] \quad . \quad . \quad . \quad (27)$$

The n.p.r. due to a noise-modulated interfering signal, where the modulation indices of both signals are very small, is therefore given by eqn. (25), degraded by additional noise in accordance with eqns. (26) and (27). The special case of $x_s = x_r$ is worth noting, although it is probably of little practical interest. The two residual carriers then interact to produce a delta-function spectral component at f_r in the baseband, as with an unmodulated interfering signal, and the n.p.r. is therefore dependent on the measuring bandwidth.

As noted in Section 2.2, the small-modulation-index approximation to the power spectrum is only valid for small values of m^2/x_1. Under these conditions, the residual carrier is large and the n.p.r. is determined substantially by eqns. (26) and (27). The interference contribution from the continuous spectra is usually comparatively negligible.

(2.4.3) **Calculation for Large Modulation Indices.**

When the modulation indices of both signals are large, eqn. (4) may be used for the mean noise power spectra. We then have, for the wanted signal,

$$f_n P_1(x) = \frac{1}{\sqrt{(2\pi) m_1}} \exp - \frac{x^2}{2m_1^2} \quad . \quad . \quad . \quad (28)$$

and, for the interfering signal,

$$f_n P_2(x) = \frac{1}{\sqrt{(2\pi) m_2'}} \exp - \frac{x^2}{2m_2'^2} \quad . \quad . \quad . \quad (29)$$

The modulation index of the interfering signal, i.e. $\Delta F_2/f_m$, where f_m is the maximum modulation frequency of the interfering signal, must be large, but the parameter m_2' need not be so.

When eqns. (28) and (29) are inserted in eqn. (21*b*), the following result is obtained for the n.p.r. at f_r in the baseband:

$$(\text{n.p.r.})_r \simeq \frac{2u^2 m_1^2}{x_r^2 (1 - x_1) I_x} \quad . \quad . \quad . \quad . \quad (30)$$

re

$$\frac{1}{2\pi m_1 m_2'}\int_{-\infty}^{\infty}\left\{\exp-\left[\frac{(x+x_a)^2}{2m_1^2}+\frac{(x-x_a)^2}{2m_2'^2}\right]\right.$$
$$\left.+\exp-\left[\frac{(x+x_b)^2}{2m_1^2}+\frac{(x-x_b)^2}{2m_2'^2}\right]\right\}dx$$

may be written

$$\frac{1}{2\pi m_1 m_2'}\left[\exp-\tfrac{1}{2}gx_a^2\int_{-\infty}^{\infty}\exp-(\tfrac{1}{2}gx^2+hx_a)dx\right.$$
$$\left.+\exp-\tfrac{1}{2}gx_b^2\int_{-\infty}^{\infty}\exp-(\tfrac{1}{2}gx^2+hx_b)dx\right]$$

re

$$g=\frac{m_1^2+m_2'^2}{m_1m_2'}$$

$$h=\frac{m_2'^2-m_1^2}{m_1m_2'}$$

e let

$$x=z-\frac{hx_a}{g}$$

ıe first integral, this becomes a probability integral of the ı

$$2\exp\frac{h^2x_a^2}{2g}\int_{-\infty}^{\infty}\exp(-\tfrac{1}{2}gz^2)dz$$

has the solution $\sqrt{(2\pi/g)}\exp(h^2x_a^2/2g)$, and a similar lt may be obtained for the second integral term in I_x, with ı place of x_a. We then have

$$=2\pi(m_1^2+m_2'^2)^{-1/2}\left[\exp-\frac{2x_a^2}{m_1^2+m_2'^2}+\exp-\frac{2x_b^2}{m_1^2+m_2'^2}\right]$$

:n this result is inserted in eqn. (30), the n.p.r. at f_r in the band due to interference from a noise-modulated signal, n the modulation indices are large, is obtained:

.r.)$_r\simeq$

$$\frac{2\sqrt{(2\pi)}u^2m_1^2(m_1^2+m_2'^2)^{1/2}/[x_r^2(1-x_1)]}{\exp-\dfrac{2x_a^2}{m_1^2+m_2'^2}+\exp-\dfrac{2x_b'}{m_1^2+m_2'^2}} \quad . \quad . \quad (31a)$$

inor manipulation yields the following:

$$.r.)_r\simeq\frac{\sqrt{(2\pi)}u^2m_1^2(m_1^2+m_2'^2)^{1/2}}{x_r^2(1-x_1)}\operatorname{sech}\frac{x_sx_r}{m_1^2+m_2'^2}$$
$$\exp\frac{x_s^2+x_r^2}{2(m_1^2+m_2'^2)} \quad . \quad . \quad . \quad . \quad . \quad . \quad (31b)$$

ay be noted from eqn. (31a) that, since m_1 is large, the effect hanges in x_r in the exponential terms is small; the factor x_r^2 ıe denominator therefore predominates, and so for all values $_s$ the n.p.r. is lowest when x_r has its maximum value of y, i.e. when $f_r=f_n$. The n.p.r. function is, in fact, very lar to that obtained for an unmodulated interfering signal. s, if m_2' is made zero in eqns. (31), eqns. (13), for the unmodu-1 interfering signal, are obtained. Any value of m_2' (implying value of $\Delta F_2/f_m$) cannot, however, be inserted in eqns. (31) btain a valid result. Apart from the special case when m_2' ero, eqns. (31) are applicable only when both noise power tra are approximately Gaussian.

the wanted and interfering signals are identical (but with orrelated modulating signals) the result given in eqns. (31) n x_s is zero may be derived by an alternative method of interest. Under these conditions, the interfering signal may be regarded as an echo signal having a very long delay, such that there is negligible correlation between the main signal and its echo. This case is discussed in Appendix 8·4. Even when the modulation indices are not large, the n.p.r. with x_s zero is clearly equivalent to what would be obtained with a very-long-delay echo. Thus, the calculated and computed results described in the present study may be applied to the problem of long-delay echo signals in f.m. systems.

(2.4.4) **Computation from Noise Spectra.**

When either the wanted or the interfering signal spectra, or both, cannot be expressed analytically, computation of the interference n.p.r. may be carried out using eqn. (21). Any of the well-known methods of numerical integration may be used to obtain the value of I_x. The integration limits are made as large as possible, using, for example, the spectrum data presented in Fig. 2. In any practical application, however, it is unnecessary to consider terms in the summation that require spectrum power densities at frequencies outside the receiver pass-band.

(3) EXPERIMENTAL INVESTIGATION

Measurements were carried out at 70 M/cs, using high-grade frequency modulators[11] and demodulators designed for use in broadband radio systems. White-noise test gear[3] was used to measure the n.p.r. due to interfering signals, after making allowance where necessary for the basic and intermodulation noise inherent in the equipment.

The noise-signal generator provided a modulating signal having a power spectrum uniform within ±0·5 dB, and extending from 60 kc/s to four selected upper frequencies: 0·55, 1·05, 2·54 and 4·03 Mc/s. Narrow band-stop filters could be inserted to produce measuring slots in the noise spectrum of at least −80 dB relative to the normal power density. The slot frequencies were 70 kc/s, 0·53 Mc/s, 1·00 Mc/s, 2·44 Mc/s and 3·89 Mc/s.

The noise receiver enabled the noise in a narrow band, selected to coincide with the slot in the generator, to be indicated on a power meter. In the measurements, only the ratio of the noise powers in the receiver band, with and without the appropriate band-stop filter inserted in the generator, was required. The variable attenuator in the receiver enabled this ratio to be measured within ±0·3 dB. Since the measuring bandwidth was relatively very small (about 1 500 c/s) the ratio measured was the required n.p.r. With the generator connected to the receiver, the limiting n.p.r. (due to noise within the receiver) was better than 75 dB in all measuring channels.

A block diagram of the equipment layout for investigating the effect of an unmodulated interfering signal is shown in Fig. 4(*a*). The unmodulated interfering signal was derived from a v.h.f. signal generator, and applied via an accurate variable attenuator to the first splitting network. By removing the noise signal from the modulator, it was possible to compare the amplitudes of the two signals on the screen of a panoramic receiver. The amplitudes were made equal, and the required discrimination against the interfering signal was obtained by adjusting the attenuation in the interfering signal path. In this way, the relative levels of the two signals could be established within about 0·3 dB.

Care was necessary to prevent spurious signals entering either the noise-signal generator or the demodulator. Each unit was well screened, and double-screened coaxial cable was used for all the connections. Also, a low-pass filter was inserted in the mains supply connections to the noise-signal generator.

Measurements were carried out to establish the variation of the n.p.r. with:

(*a*) Variation of the interfering signal level relative to the wanted carrier level, over a range of zero to −50 dB.

(*b*) Variation of frequency separation over the range zero to ±30 Mc/s.
(*c*) Variation of the r.m.s. modulation index over the range 0·04 to 2.
(*d*) Variation of the frequency of the measuring channel in the baseband.

In the investigation of a noise-modulated interfering signal, a similar method was adopted, but a rather more elaborate equipment layout was necessary, and this is shown in Fig. 4(*b*). The main differences in the layout are

(i) A second, similar frequency modulator was used as the interfering signal source.
(ii) A second, similar noise-signal generator was provided to modulate the interfering signal.

The two signals were again displayed on the screen of the panoramic receiver so that the unmodulated carrier amplitudes could be compared. A third, unmodulated, signal was also applied initially from the v.h.f. signal generator to the panoramic receiver, to check and set the frequency of the interfering signal.

A similar programme of measurements was undertaken as for the unmodulated interfering signal. The maximum frequency separation of the mean carriers that could be achieved was, however, restricted to ±10 Mc/s, but this proved to be adequate. The same deviation and modulating signal band were used for the two f.m. signals.

The limiters in the demodulator used were designed to provi a high degree of dynamic amplitude limiting with minimu conversion of amplitude to phase modulation. The performan of the limiters was examined in the following way. An unmod lated interfering signal, together with an unmodulated want carrier signal, was applied to the demodulator, with the inte fering signal level 20 dB below that of the wanted signal. T output of the demodulator was applied to the noise receiver a a measuring channel selected. It was then possible to adju the frequencies of the two signals to obtain a baseband sign in the measuring channel, and to measure the effective frequen deviation imparted to the wanted carrier. This was carried o in all the measuring channels, and at various signal frequenci in the demodulator pass-band; in all cases, the measured devi tions agreed with the calculated values within the measuri error of about 5%. If the limiting had been inadequate, c amplitude modulation had been converted to phase modulati this result would not have been obtained. The effect of ina quate limiting is considered by Medhurst *et al.*, and it may seen from eqn. (9) of Reference 2 that, if the amplitude compr sion factor, Δ_L, is very small, the effect discussed is insignific In the demodulator used in the present work this factor is le than 0·005.

The overall limiting n.p.r. obtained with the equipm described, in the absence of an interfering signal, varied betw

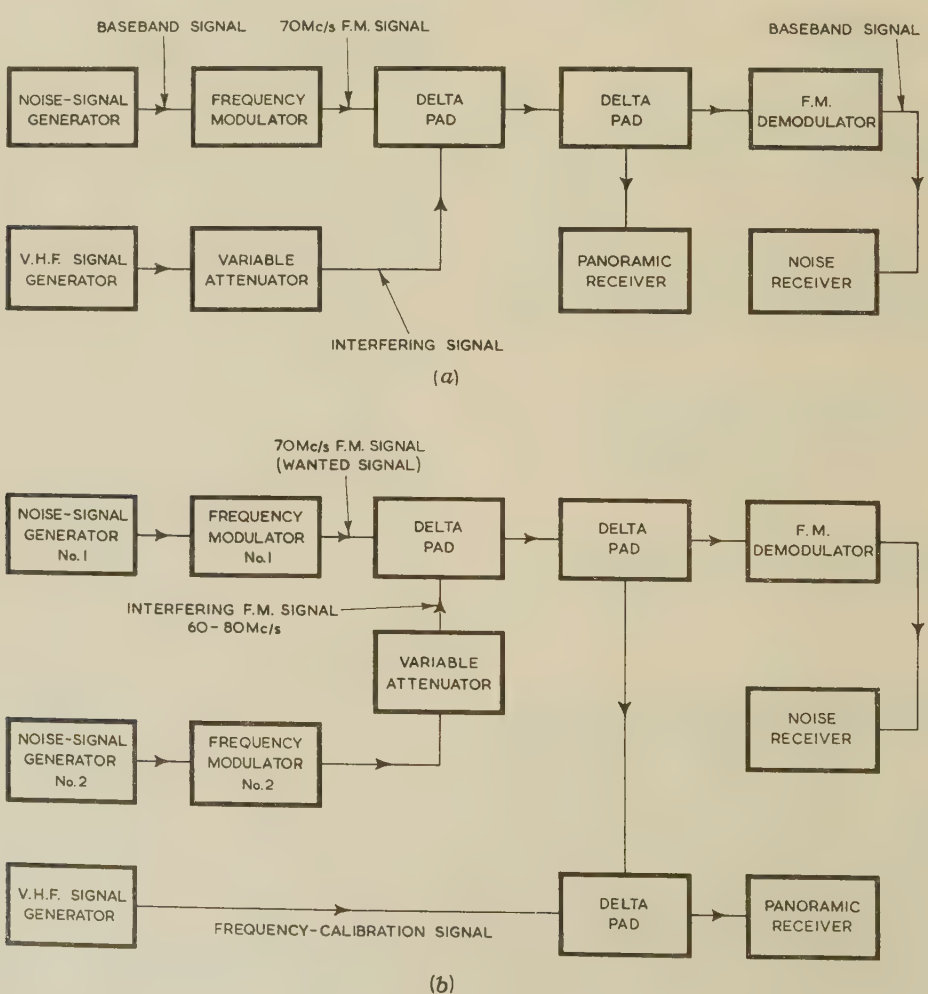


Fig. 4.—Measuring equipment.
(*a*) Unmodulated interfering signal.
(*b*) Noise-modulated interfering signal.

ɔut 50 and 65 dB, depending on the modulation characteristics. e maximum n.p.r. that could be measured did not, therefore, eed these limiting values by more than a few decibels, and the or in the final n.p.r. readings was not expected to exceed dB.

(4) RESULTS OF MEASUREMENTS

(4.1) Unmodulated Interfering Signal

A preliminary investigation confirmed that the n.p.r. in any asuring channel is proportional to u^2, where u is the ratio wanted to interfering signal amplitudes. This relation was cked over a range of r.m.s. modulation indices, m, of 0·04–2, rl at various frequency separations. For ratios between 5 and dB (the largest value used) the relationship held to within than 1 dB. The result was found to be valid within 2 dB h ratios as low as 1 dB, but for smaller values the limiting ion in the demodulator failed, causing the n.p.r. to be essive. It may be concluded that the theoretical relationship been confirmed, as concluded also in Reference 2. In the nainder of the measurements an amplitude ratio of 20 dB = 10) was used, except in a few cases where a value of 10 dB s necessary to bring the measured n.p.r. below the limiting ue.

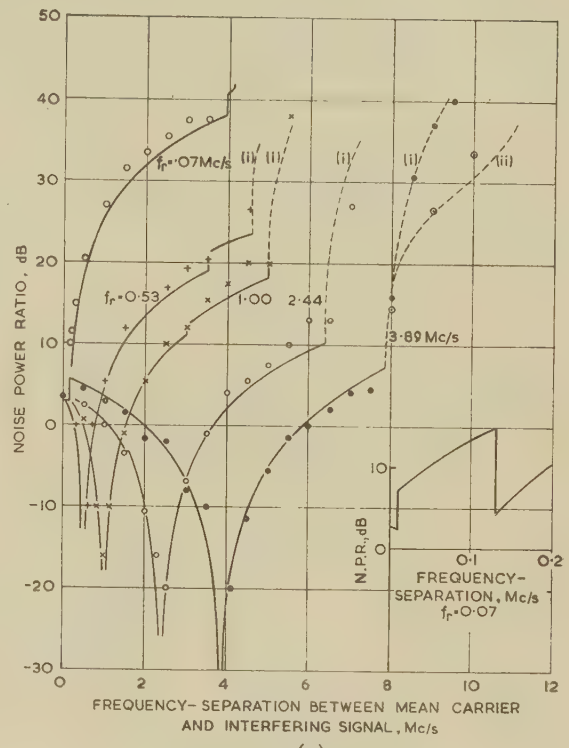
The results of the measurements relating the n.p.r. to the quency separation, for various modulation indices and at eral baseband frequencies, are shown in Figs. 5–7. Theoretical ves have been calculated or computed for the cases examined, l the measured values are plotted for comparison with the ves. Each value plotted is a power mean (arithmetic mean the reciprocal n.p.r.'s) of about six measurements. The lts are shown normalized for $u = 1$, and only positive values frequency separation are shown; however, the measurements luded interfering signal frequencies above and below the ated carrier frequency.

The results for very small modulation indices are shown in . 5, where the main portions of the theoretical curves have n calculated (solid lines). The tails have been computed m the spectrum curves of Fig. 2. It may be seen that the asured values (for $f_r = 3{\cdot}89$ Mc/s) confirm that for very all modulation indices the n.p.r. is independent of the index. expected, the measured n.p.r. when $f_s = f_r$ was very low, could not be reliably measured and has not been calculated. e detail near $f_s = 0$ for $f_r = 70$ kc/s is shown in Fig. 5(*a*) but uld not be resolved by measurement.

n Fig. 6(*a*), the curve for $m = 0{\cdot}2$ was computed ($m^2/x_1 =$ 5); in the measurement, the value of m^2/x_1 was 1·7, but od agreement is shown. This might be expected, since the ct of the lowest modulation frequency, and hence of m^2/x_1, omes progressively less as m increases. In curve (ii), for = 0·28, and in the succeeding curves for larger values of m, results are substantially independent of x_1.

n Fig. 6(*b*) ($m = 0{\cdot}95$) the measured values for $f_r = 1{\cdot}0$ Mc/s compared with both a calculated curve, assuming a Gaussian ctrum, and a curve computed from the measured spectrum. nay be noted that the calculated curve is seriously in error. nen $m = 2$, however, it may be seen from Fig. 7 that the asured values were in good agreement with the curve calculated uming a Gaussian spectrum.

Measurements were made up to the maximum frequency aration at which n.p.r. readings could be obtained. In lition, however, the n.p.r. was observed as the frequency aration was increased up to ± 30 Mc/s. There were a few uctions in n.p.r. at certain frequencies of the interfering signal, , except over the regions plotted, the n.p.r. remained at the iting value over this range of frequency separation. The



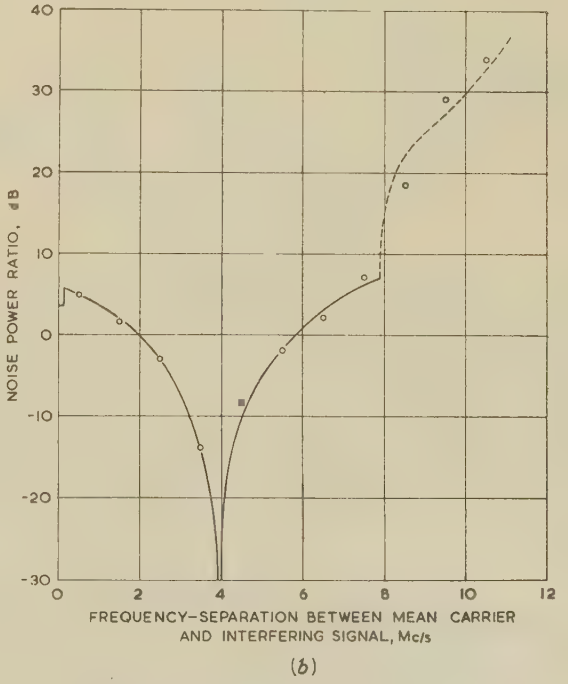
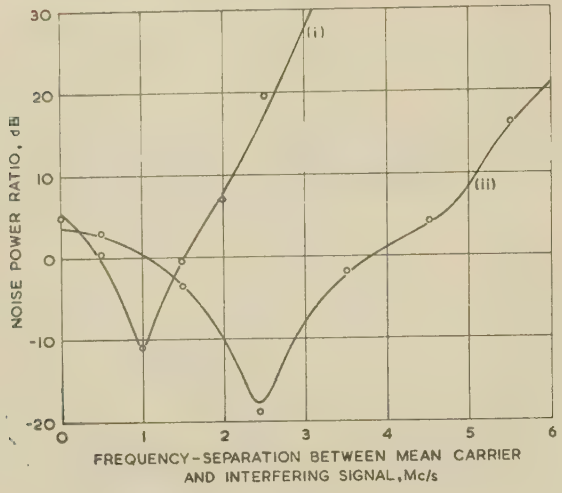


Fig. 5.—Unmodulated interfering signal.

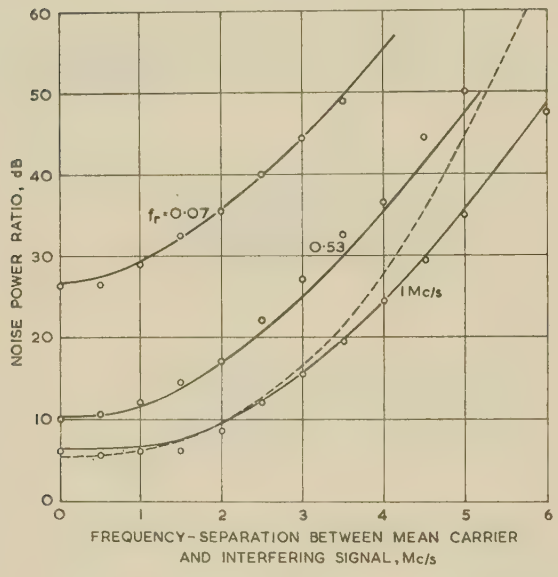
N.P.R., when modulation index is very small, normalized for $u = 1$.

(*a*) Solid curves calculated ($m \ll 1$); broken portions computed with $m \leqslant 0{\cdot}1$ and (i) $m^2/x_1 = 0{\cdot}1$, (ii) $m^2/x_1 = 0{\cdot}6$. Measured values plotted ($m = 0{\cdot}04$, $m^2/x_1 = 0{\cdot}11$, but for points ⊙ $m^2/x_1 = 0{\cdot}26$).

(*b*) Solid portion calculated ($m \ll 1$); broken portion computed ($m = 0{\cdot}1$, $m^2/x = 0{\cdot}6$). Measured values plotted ($m = 0{\cdot}12$, $m^2/x_1 = 1$).



(*a*)



(*b*)

Fig. 6.—Unmodulated interfering signal.

N.P.R. when modulation index is small, normalized for $u = 1$.
(*a*) Curves computed for (*i*) $m = 0{\cdot}2$ and (*ii*) $m = 0{\cdot}3$. Measured valves plotted ($m = 0{\cdot}2$ and $0{\cdot}28$).
(*b*) Solid curves computed for $m = 1$.
Broken curve calculated assuming Gaussian spectrum.
Measured values plotted ($m = 0{\cdot}95$).

abnormal reductions were identified as arising from residual interference from a local television transmitter. Apart from this particular effect, no evidence was observed of any limiting value of n.p.r. at large frequency separations as noted in Reference 2.

(4.2) Noise-Modulated Interfering Signal

A preliminary investigation again showed that, with a noise-modulated interfering signal, the n.p.r. is always proportional to u^2, for u greater than about 1·5.

The principal results of the measurements using an interfering signal with the same modulation characteristics as the wanted signal are presented in Figs. 8–10.

The measured values of n.p.r. are compared with calculated curves, for very small modulation indices, in Fig. 8. The tails of the f.m. spectra have not been included in the calculation, so

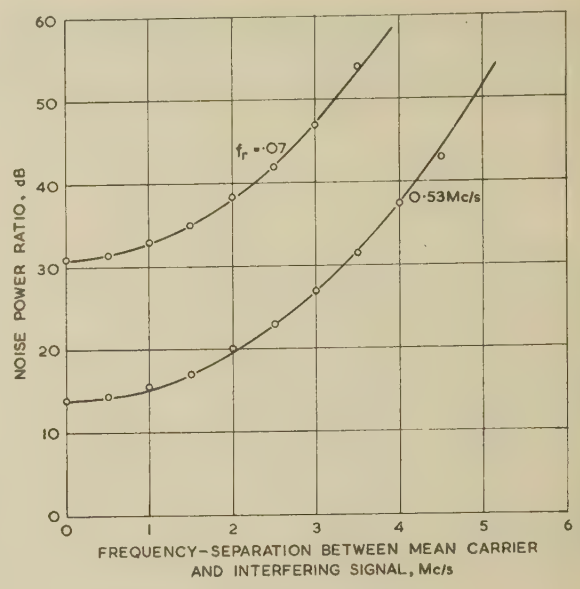


Fig. 7.—Unmodulated interfering signal.

N.P.R. when modulation index is large, normalized for $u = 1$.
Curves calculated for $m = 2$ assuming Gaussian spectrum.
Measured values plotted ($m = 2$).

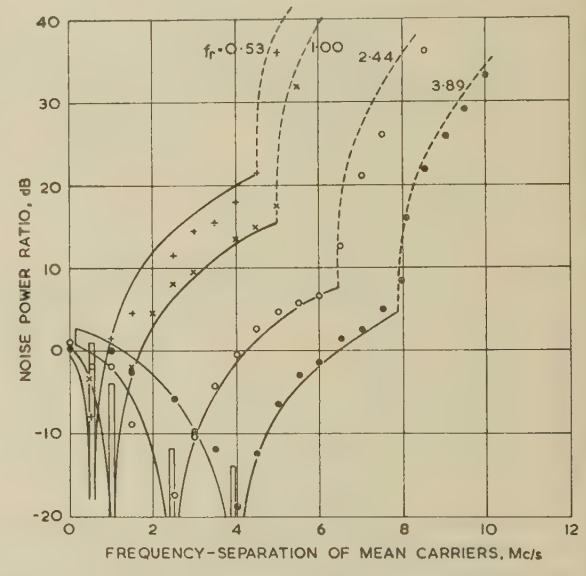


Fig. 8.—Noise-modulated interfering signal.

N.P.R. when both modulation indices are very small, normalized for $u = 1$.
Solid curves calculated ($m \ll 1$) and broken portions computed for $m = 0{\cdot}6$, $m^2/x_1 = 0{\cdot}6$.
Measured values plotted ($m = 0{\cdot}04$, $m^2/x_1 = 0{\cdot}1$).

the measured values were expected to fall below the curves at large frequency separations. The agreement is quite good apart from the results for $f_r = 0{\cdot}53$ Mc/s. Also, consistent measurements could not be obtained with $f_r = 0{\cdot}07$ Mc/s. The reasons for these discrepancies are not known, but it was established that they were not due to inadequate limiting in the demodulator. The effect of the residual carriers interacting together is omitted.

The results for $m = 0{\cdot}28$, at one baseband frequency, are shown in Fig. 9.

In Fig. 10, the results at two baseband frequencies are shown with $m = 0{\cdot}95$. The measurements with $f_r = 1{\cdot}0$ Mc/s are compared with both a theoretical curve based on Gaussian spectra and a curve computed from the measured spectrum

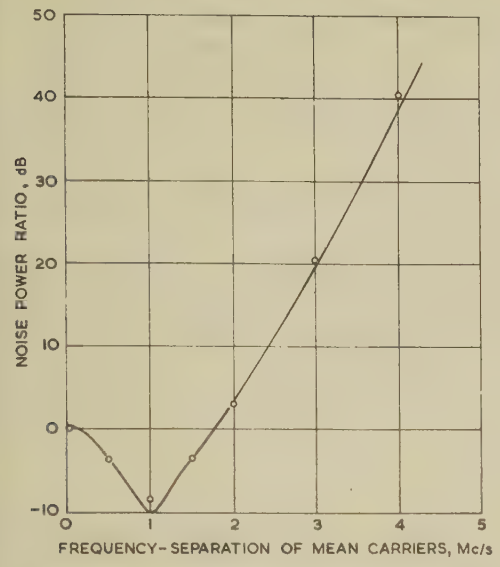


Fig. 9.—Noise-modulated interfering signal.

N.P.R. when modulation indices are small, normalized for $u = 1$.
Curve computed for $m = 0{\cdot}3$.
Measured values plotted ($m = 0{\cdot}28$).

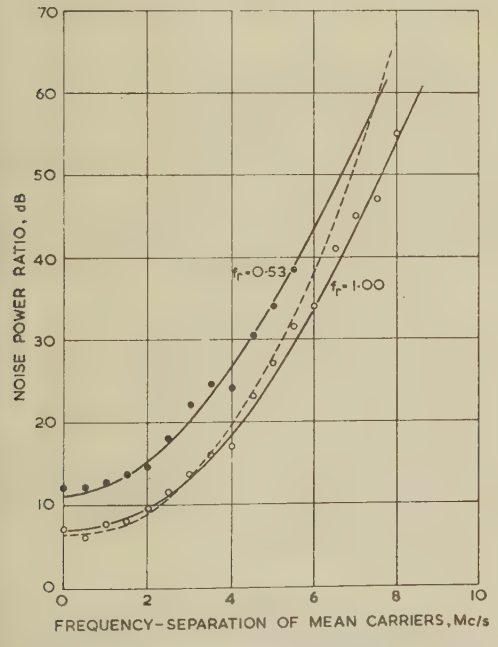


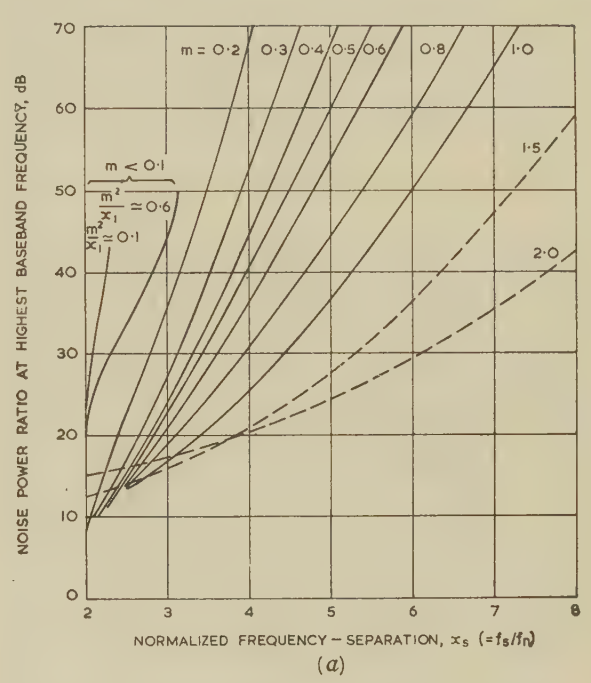
Fig. 10.—Noise-modulated interfering signal.

P.R. when modulation indices are large, normalized for $u = 1$.
id curves computed ($m = 1$) and broken curve calculated assuming Gaussian
a.
asured values plotted ($m = 1$).

measurements are seen to follow more closely the computed
e. It was again difficult to obtain consistent measurements at
seband frequency of 0·07 Mc/s.
he measured results for $m = 2$ were compared with a cal-
ted curve, assuming Gaussian spectra, and good agreement
found.

(5) CONCLUSIONS AND GENERAL RESULTS

/ith the exception of a few minor discrepancies, believed to
n from equipment limitations, the results presented in Sec-
. 4 are considered to confirm the validity of the theoretical analysis. Moreover, the theory is found to be in reasonably good agreement with the measurements for very small modulation indices, where quasi-stationary conditions do not apply. Consequently, it may be concluded that the initial hypothesis discussed in Section 2.1 is valid.



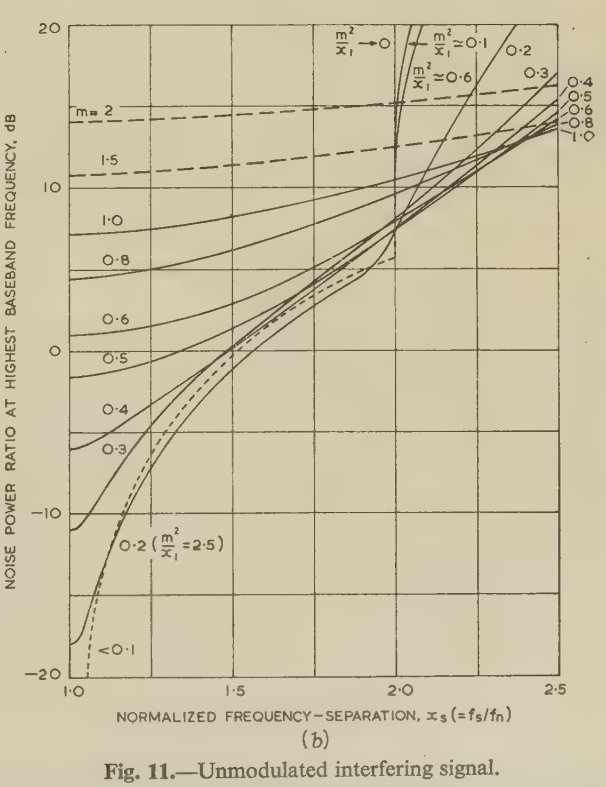


Fig. 11.—Unmodulated interfering signal.

Generalized results of N.P.R. against frequency separation.
Broken curves calculated, solid curves computed, for 10 values of m.
For actual n.p.r. add 20 log u decibels, where u = (wanted carrier amplitude)/(interfering signal amplitude).
(*a*) Curves for normalized frequency separations of 2–8.
(*b*) Detail for small frequency separation.

In order to enhance the value of the results in practical applications, general curves relating the n.p.r. to the frequency separation have been prepared. These are shown in Fig. 11 for an unmodulated interfering signal and in Fig. 12 for a noise-modulated interfering signal. The results are normalized for $u = 1$, and are necessarily given in terms of the reduc variables. The n.p.r. at the highest baseband frequency o is given, since this is of greatest interest in practice. In Fig. the curves relate to two signals having identical modulati characteristics, since this is the most usual situation encounte in practical problems. The results of the theoretical analy are, of course, available, in conjunction with the normali noise power spectra in Fig. 2, for the solution of particu problems in this field.

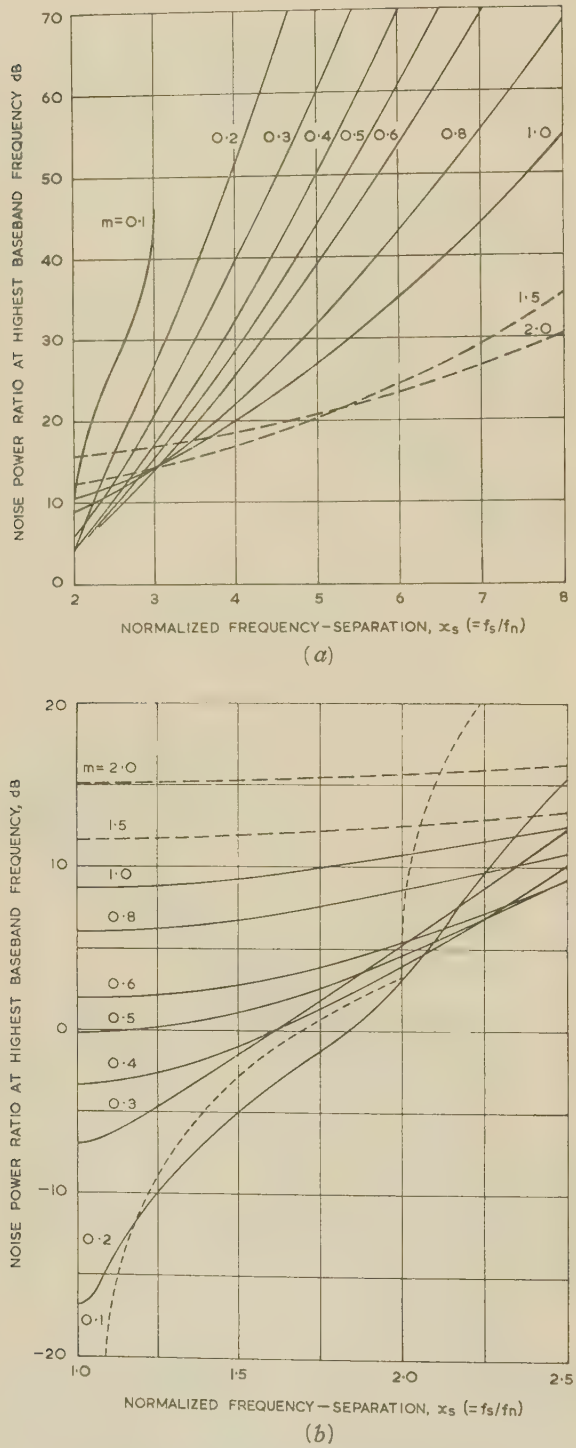


Fig. 12.—Noise-modulated interfering signal.

Generalized results of n.p.r. against frequency separation.
Broken curves calculated, solid curves computed, for 10 values of m.
For actual n.p.r. add 20 log u decibels, where u = (wanted carrier amplitude)/(interfering carrier amplitude).

(*a*) Curves for normalized frequency separations of 2–8.
(*b*) Detail for small frequency separation.

It may be concluded from the theoretical analyses that, two or more uncorrelated interfering signals of the types c sidered are present, the n.p.r. can be deduced by adding interference power densities from the several sources. T conclusion is relevant in the choice of frequency separat between the carriers of f.m. radio systems operating in a com frequency band. In this case, the sum of the amplitudes o interfering signals must be small compared with that of wanted signal.

The hypothesis that the normalized power density of an f. signal is equal to the probability density of the instantan frequency requires further comment. Although intuit plausible, it is in apparent conflict with the general represent of an f.m. signal in the time domain. Nevertheless, the exp sion for the instantaneous frequency of an f.m. signal limited value analytically, whereas the present work shows the suggested hypothesis is useful when dealing with both frequency and time domains. It may be concluded that it validity in the interpretation of f.m. power spectra in the domain. Moreover, it may also be concluded that the hy thesis is valid when applied to line spectra; i.e. the probabi that the instantaneous frequency of an f.m. signal coinci with a particular spectrum component is equal to the normali power of that component.

The extension of the method described to phase-modula systems presents little difficulty; use may be made of the pha modulation noise power spectra described in Reference 5.

(6) ACKNOWLEDGMENTS

The author wishes to thank the Engineer-in-Chief of the P Office, and the Director of Research in the Post Office Engine ing Department, for permission to publish the paper.

The encouragement and interest shown by Mr. R. W. Wh and the assistance of the laboratory staff in constructing m of the equipment described, are gratefully acknowledged.

The paper is published by permission of the Controller Her Majesty's Stationery Office.

(7) REFERENCES

(1) 'Extraneous R.F. Interference in F.M. Multicha Systems': Post Office Radio Report (describing w carried out in 1955).

(2) MEDHURST, R. G., HICKS, E. M., and GROSSETT, 'Distortion in Frequency-Division-Multiplex F Systems due to an Interfering Carrier', *Proceedings I.E* Paper No. 2565 R, May, 1958 (**105** B, p. 282).

(3) WHITE, R. W., and WHYTE, J. S.: 'Equipment for Meas ment of Inter-Channel Crosstalk and Noise on Bro Band Multichannel Telephone Systems', *Post Of Electrical Engineers' Journal*, 1955, **48**, p. 127.

(4) GLADWIN, A. S.: 'Energy Distribution in the Spectrum a Frequency-Modulated Wave', *Philosophical Magaz* 1944, **35**, p. 787; *ibid.*, 1947, **38**, p. 229.

(5) HAMER, R., and ACTON, R. A.: 'Power Spectrum o Carrier Modulated in Phase or Frequency by W Noise', *Electronic and Radio Engineer*, 1957, **34**, p. 24

VAN DER POL, B.: 'The Fundamental Principles of Frequency Modulation', *Journal I.E.E.*, 1945, **93**, Part III, p. 153.

MIDDLETON, D.: 'The Distribution of Energy in Randomly Modulated Waves', *Philosophical Magazine*, 1951, **42**, p. 689.

STEWART, J. L.: 'The Power Spectrum of a Carrier Frequency Modulated by Gaussian Noise', *Proceedings of the Institute of Radio Engineers*, 1954, **42**, p. 1939.

BLACHMAN, N. M.: 'Limiting Frequency-Modulation Spectra', *Information and Control*, 1957, **1**, p. 26.

CHERRY, E. C., and RIVLIN, R. S.: 'Non-linear Distortion, with particular reference to the Theory of Frequency-Modulated Waves', *Philosophical Magazine*, 1941, **32**, p. 265.

RAVENSCROFT, I. A.: 'An Improved Frequency Modulator for Broad-Band Radio-Relay Systems', *Post Office Electrical Engineers' Journal*, 1957, **50**, p. 186.

LEWIN, L.: 'Interference in Multi-Channel Circuits', *Wireless Engineer*, 1950, **27**, p. 294.

BENNETT, W. R., CURTIS, H. E., and RICE, S. O.: 'Interchannel Interference in F.M. and P.M. Systems under Noise Loading Conditions', *Bell System Technical Journal*, 1955, **34**, p. 601.

RICE, S. O.: 'Mathematical Analysis of Random Noise: Part III, Statistical Properties of Noise Current', *ibid.*, 1945, **24**, p. 46.

(8) APPENDICES

(8.1) Intelligible Interference

It is shown in Reference 2 that, when the mean frequency aration of the wanted and interfering carriers is less than ce the maximum modulation frequency of the wanted signal, orm of intelligible interference is produced. Using the ıbols of the present work, it is shown that the n.p.r. in a small ıd centred on a frequency f_r in the baseband, due to intelble interference caused by an unmodulated interfering ıal, is

$$(\text{n.p.r.})_r \simeq 4u^2\frac{x_q^2}{x_r^2}\exp\frac{m^2}{x_1} \quad . \quad . \quad . \quad . \quad (32)$$

ere $x_q = f_q/f_n$

f_q = Mean frequency in the baseband of a small band from which intelligible interference is derived.

The condition for this type of interference to occur is shown Reference 2 to be $x_r = x_s + x_q$ or $x_r = |x_s - x_q|$. This plies that a signal at f_q in the baseband is transferred, at a luced amplitude, to f_r in the baseband. The interference ould, perhaps, be termed quasi-intelligible, for the following sons:

(*a*) There are two possible baseband frequencies, f_q and f_q', from vhich interference is derived and, in one case, frequency inversion ccurs; i.e. a small band at f_q, say (representing a telephone channel n the f.d.m. system) is inverted when transferred to a band at f_r.

(*b*) In practice the relative frequency of the interfering signal, f_s and hence x_s) will not remain precisely fixed, a comparatively ninute change in frequency causing a different telephone channel o replace the one originally at f_q.

Eqn. (32) may be derived using the arguments following. nsider a single f.m. first-order side frequency of normalized plitude v in the band δf at A [Fig. 1(*a*)] of relative frequency f_q. may then be supposed that the f.m. signal resides at this side quency for a proportion of the time equal to v^2, assuming it carrier power and voltage. Thus, if the ratio of wanted to interfering signal amplitudes is u ($\geqslant 1$), the mean interference power, P_i, at f_r in the baseband, due to the component v, is

$$P_i \propto \frac{f_r^2}{2u^2}v^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (33)$$

This is not the total interference power produced, but only that due to the single first-order sideband component considered. Considering now the effect of this component (and its image) within the f.m. spectrum, it is readily deduced that the phase modulation imparted to the residual carrier produces, at f_q in the baseband, a mean wanted signal power, P_s, given by

$$P_s \propto \frac{2f_q^2}{P_D}v^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (34)$$

where P_D is the residual carrier power. However, since we are dealing with a modulating signal of uniform power density, the wanted signal power at f_r in the baseband, due to a similar component to that considered at f_q, is also given by eqn. (34). Thus, the ratio of the mean wanted to interfering signal power in a band δf in the baseband, from the cause considered, is P_s/P_i, since the same arguments may be applied to each of any number of side-frequency components of the type examined. The constant of proportionality is the same in eqns. (33) and (34), so eqn. (32) is readily established.

This result appears to be valid only when the modulation index is very small, for, apart from this initial assumption, only the wanted signal power corresponding to the first-order f.m. spectrum has been considered. However, it is likely to be a good approximation at all values of modulation index, since the magnitude of the exponential factor is itself a measure of the extent to which the assumptions are valid. It follows from eqn. (32) that the larger the value of m^2/x_1, and, in general, the larger m, the smaller will be the intelligible part of the interference.

When the interfering signal is also frequency modulated by a noise signal, quasi-intelligible interference is produced in the manner described, regarding the residual carrier of the interfering signal as an unmodulated signal. Thus, this type of intelligible interference is negligible unless the residual carrier of the interfering signal is relatively large. The n.p.r. from this cause is readily shown to be

$$(\text{n.p.r.})_r \simeq 4u^2\frac{x_q^2}{x_r^2}\exp\left(\frac{m_1^2}{x_1}+\frac{m_2^2}{x_2}\right) \quad . \quad . \quad . \quad (35)$$

If the interaction of the interfering signal spectrum with the residual carrier of the wanted signal is considered, a similar result to that of eqn. (35) is obtained. However, the quasi-intelligible interference is then from the baseband of the interfering signal into that of the wanted signal.

Thus, if the modulation index of either f.m. signal, or of both, is small, significant quasi-intelligible interference (or crosstalk between the systems) is theoretically possible. For moderate and large modulation indices the effect is insignificant. Also, as already noted, and pointed out in Reference 2, the effect is unimportant in practice owing to unavoidable frequency instabilities. It may be noted that the values of n.p.r. deduced in the present work include the effect of quasi-intelligible interference.

(8.2) Effect of an Interfering Signal comparable in Amplitude with the Wanted Signal

When the ratio, u, of the wanted to interfering signal amplitudes is large, the approximation used in Section 2.3.1 is permissible. It may be shown that if $u > 3$, the error is less than 2%, including the assumption that $V_s = V_r$ (see Fig. 3).

When the amplitude of the interfering signal is comparable

with that of the wanted signal, a better approximation may be obtained in the following way. From the vector diagram, we have

$$\phi = \arctan\left(\frac{\sin \omega_r t}{u + \cos \omega_r t}\right)$$

where $2\pi f_r$ has been replaced by ω_r. On expanding in a series, we have

$$\phi \simeq \frac{1}{u}\sin \omega_r t - \frac{1}{2u^2}\sin 2\omega_r t + \frac{1}{3u^3}\sin 3\omega_r t + \ldots$$

As would be expected, the spurious frequency modulation consists of a series of terms, harmonically related in frequency, only the first of which is important when u is large. Considering an unmodulated interfering signal, it may be seen that spurious frequency modulation arises at a frequency f_r from all the terms, but due to different instantaneous frequency separations. For the three terms shown, interference arises from three pairs of bands, δf, in the f.m. spectrum, instead of just the one pair dealt with in Section 2.3. These bands are separated from the interfering signal by frequencies of f_r, $\frac{1}{2}f_r$ and $\frac{1}{3}f_r$. The resultant spurious r.m.s. frequency deviation, ΔF_i, for the three terms is thus (noting that the second-harmonic power falling at f_r derives from a small band $\frac{1}{2}\delta f$, the third from a band $\frac{1}{3}\delta f$, and so on)

$$\Delta f_i^2 \simeq \frac{f_r^2}{2u^2} + \frac{f_r^2}{16u^4} + \frac{f_r^2}{54u^6} \quad . \quad . \quad . \quad . \quad (36)$$

Consequently, using this result and the appropriate analysis as in Section 2.3.1, eqn. (10) becomes

$$(\text{n.p.r.})_r \simeq \frac{m^2}{x_r^2(1-x_1)}\Big/\left\{\frac{1}{2u^2}[f_n P(x_s - x_r) + f_n P(x_s + x_r)]\right.$$
$$+ \frac{1}{16u^4}[f_n P(x_s - \tfrac{1}{2}x_r) + f_n P(x_s + \tfrac{1}{2}x_r)]$$
$$\left. + \frac{1}{54u^6}[f_n P(x_s - \tfrac{1}{3}x_r) + f_n P(x_s + \tfrac{1}{3}x_r)]\right\}$$

The same arguments apply in the case of a noise-modulated interfering signal, and a similar result may readily be derived to replace eqn. (21). In practice, it is unlikely to be necessary to consider more than the first term, since u will generally be sufficiently large.

(8.3) Interference Probabilities

With a noise-modulated interfering signal, two small spectrum bands, δf, spaced f_r apart, give rise to interference in a band $f_r \pm \delta f$ in the baseband. These two bands are shown in Fig. 13, in which power densities (or interference probability densities) of p_1 and p_2 are assumed. However, as has been noted, if one carrier occupies the band at A, there are two partial-probability bands, δf wide, adjoining the band at B, which also give rise to interference in the band $f_r \pm \delta f$ in the baseband. The effect of these bands is examined here.

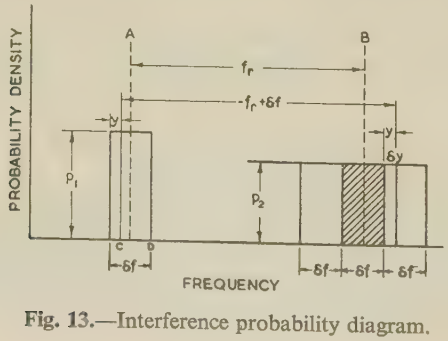


Fig. 13.—Interference probability diagram.

One carrier occupies the band at A and the other simultaneously the cross-hatched band at B with a probability

$$\mathscr{P}_1\mathscr{P}_2 = p_1\delta f \,.\, p_2\delta f$$

Now suppose that whilst the first carrier is in the band at A the second carrier occupies a relatively infinitesimal band, δy spaced $(\frac{1}{2}\delta f + y)$ from the centre of the cross-hatched band at B as shown in the diagram. Interference can now fall in the band $f_r \pm \delta f$ in the baseband only if the first carrier occupies the portion CD of the band at A. The probability, $\mathscr{P}$, of this occurring is

$$\mathscr{P} = p_1\text{CD} \,.\, p_2\delta y$$

But, from the diagram, $\text{CD} = \delta f - y$. Therefore, the total required probability of interference, whilst the second carrier occupies the right-hand adjoining band at B, is

$$\mathscr{P} = p_1p_2\int_0^{\delta f}(\delta f - y)dy$$
$$= \tfrac{1}{2}p_1\delta f \,.\, p_2\delta f$$

That is,
$$\mathscr{P} = \tfrac{1}{2}\mathscr{P}_1\mathscr{P}_2$$

The same arguments apply when the second carrier occupies the left-hand adjoining band at B, the probability again being $\frac{1}{2}\mathscr{P}_1\mathscr{P}_2$. Thus, the probability taking into account both partial-probability bands is $\mathscr{P}_1\mathscr{P}_2$, and the total probability, including the cross-hatched full-probability band at B, is therefore $2\mathscr{P}_1\mathscr{P}_2$. When the inverse positioning of the wanted and interfering carriers is considered, a similar result is obtained; the contribution of the partial-probability bands at A is again equal to that of the full-probability band at A.

(8.4) Analogy with a Very-Long-Delay Echo Signal

In f.m. multi-channel telephony transmission systems, echo signals cause intermodulation distortion of the baseband signal. Such echo signals arise owing to feeder reflections and multi-path radio propagation, and are usually of short delay. The effect of echo signals has been studied[12, 13] with the emphasis chiefly on the important cases where the delay is small compared with the period or the maximum-frequency modulating signal. In future waveguide transmission systems, however, the case of a very-long-delay echo signal may be important.

In the limit, a long-delay echo signal may be regarded as a similarly modulated interfering signal with the same mean carrier frequency but with an uncorrelated modulating signal. Thus, if the multi-channel modulating signal is represented by a random-noise signal, the n.p.r. resulting with a noise-modulated interfering signal of the same mean frequency should apply also to a long-delay echo signal. The results given in Section 3 with f_s zero, should therefore apply. This has been checked experimentally, using a 70 Mc/s signal frequency-modulated by noise signal, with long coaxial cables to provide an echo signal. It was found that the n.p.r. was asymptotic to the value expected with good agreement for values of the parameter $2\pi f_n T$ exceeding about π; here f_n is the highest modulation frequency and T the echo delay.

When the modulation index is large, the n.p.r. associated with a long-delay echo signal may be derived by a method closely analogous to that used in Section 2.4.1. The frequency/time distribution of the carrier is in this case Gaussian, and a result given by Rice[14] for the joint probability density of two super-

osed random functions, $s(t)$ and $s(t+T)$, is applicable; is

$$= (\psi_0^2 - \psi_T^2)^{-1/2} \frac{\delta f^2}{2\pi} \exp - \left(\frac{\psi_0 f_d^2 + \psi_0 f_e^2 - 2\psi_T f_d f_e)}{2(\psi_0^2 - \psi_T^2)}\right). \quad (37)$$

ere $\mathscr{P}$ = Probability that $s(t)$ lies between f_d and $(f_d + \delta f)$ at the same time as $s(t+T)$ lies between f_e and $(f_e + \delta f)$.

ψ_T = Autocorrelation function of $s(t)$.

$\psi_0 = \psi_T$ with $T = 0$; i.e. mean-square value of $s(t)$.

he case being considered, $s(t)$ is a frequency function of time; ce ψ_0 may be identified as the mean-square deviation, ΔF^2. ve stipulate a uniform-spectrum noise signal extending from o to f_n we have

$$\psi_T = \Delta F^2 \frac{\sin\beta}{\beta}$$

ere β is $2\pi f_n T$. Referring to Fig. 1(*b*), it may be seen that, spurious frequency modulation to be produced at a fre- ncy f_r,

$$f_d = f - \tfrac{1}{2} f_r$$
$$f_e = f + \tfrac{1}{2} f_r$$

ere f is as defined with reference to the diagram. Inserting se values in eqn. (37), we have

$$f) = \frac{2\delta f^2}{\pi\left\{\Delta F^4\left[1 - \left(\frac{\sin\beta}{\beta}\right)^2\right]\right\}^{1/2}}$$

$$\exp - \frac{\Delta F^2 f^2 + \tfrac{1}{4}\Delta F^2 f_r^2 - (f^2 - \tfrac{1}{4} f_r^2)\Delta F^2 \frac{\sin\beta}{\beta}}{\Delta F^4\left[1 - \left(\frac{\sin\beta}{\beta}\right)^2\right]}$$

ting that a factor of 4 has been applied to take account of the erse band occupancy and the partial-probability bands, as discussed in Section 2.4.1. The function $\mathscr{P}(f)$ is here identified as the interference probability, and the full mean probability is obtained by integration; thus, we have

$$\mathscr{P} = \frac{2\delta f}{\sqrt{\pi}\Delta F\left(1 - \frac{\sin\beta}{\beta}\right)^{1/2}} \exp \frac{f_r^2}{4\Delta F^2\left(1 - \frac{\sin\beta}{\beta}\right)}$$

The reciprocal of the n.p.r. is the product of $\mathscr{P}$ and the instantaneous interference to mean-signal power ratio, P_i'/P_s, given in eqn. (14), for here $\Delta F_1 = \Delta F_2 = \Delta F$. The required result is therefore

$$\text{(n.p.r.)} \simeq \frac{2\sqrt{\pi} u^2 \Delta F^3\left(1 - \frac{\sin\beta}{\beta}\right)^{1/2}}{f_r^2 f_n} \exp \frac{f_r^2}{4\Delta F^2\left(1 - \frac{\sin\beta}{\beta}\right)} \quad . \quad (38)$$

where f_1 has been made zero.

Eqn. (38) gives the n.p.r. resulting from an echo signal in an f.m. system in which the modulation index is large. No restriction has yet been placed on the echo delay, but the result is not of general validity, since no account has been taken of the phase correlation of the carrier and its echo. It has been found experimentally that eqn. (38) is a good approximation with values of β exceeding about $\pi/4$, the average n.p.r. for smaller values of β being lower than is given by the equation.

When the echo delay is very long, the phase correlation of the carrier and its echo becomes negligible, and eqn. (38) is a valid approximation. In this case, also, as $T \to \infty$, $(\sin\beta)/\beta \to 0$, and the equation is further simplified; using the reduced variables, the result is

$$(\text{n.p.r.})_r = \frac{2\sqrt{\pi} u^2 m^3}{x_r^2} \exp \frac{x_r^2}{4m^2}$$

This is the result obtained by putting $x_s = 0$ (and $x_1 = 0$), and $m_1 = m_2 = m$ in eqn. (31*b*); it corresponds to case 6(*d*) of Reference 13 (p. 623).

621.385.032.213.13

The Institution of Electrical Engineers
Paper No. 3333 E
Jan. 196

EMISSION FROM MINIATURE HOLLOW CATHODES

By A. SANDOR, M.E.E., D.E.Sc.

(*The paper was first received 2nd October, 1959, in revised form 24th March, and in final form 30th June, 1960.*)

SUMMARY

Miniature hollow cathodes, internally coated with conventional oxides, have been studied. As visual inspection revealed, oxide deposits at the narrow orifice edge constricted the flow of electrons, and the initial emission from this edge was then spontaneously destroyed by saturation-caused deactivation, owing to positive-ion action. Large hexagonal crystal growth on the outer nickel body was also observed. While the external field causes the electrons to emerge from the cavity as a laminar flow, the conditions in the cavity change abruptly to a gas discharge with large field, and at this point ion bombardment damages the coating. The emission characteristics suggest a strong temperature dependence of the electron flow to an external collector. Although a current density of 10 amp/cm^2 at 965° C was obtained where the proximity of the anode worked against early ionization breakdown, a load of 7 amp/cm^2 at 920° C was maintained in other cases for 800 hours. Modified orifice arrangements showed that optimum emission is obtained with a sharp orifice edge and solid cavity walls, although no difference in emission was detected between a fully and a partly coated cavity. Grid-controlled operation revealed a decrease in transconductance, owing to field shielding and to the receding virtual cathode, while the focusing quality suffered mainly by the added orifice lens.

In Section 3 an interpretation of the mechanisms of current flow to an internal and an external anode is attempted. Since the emission characteristics do not reveal the presence of the underlying emission laws in an explicit form, emission is related to an equivalent diode with dynamic perveance. Hollowness in the beam cross-section at higher currents is interpreted as the result of bunching near the orifice edge, localized current saturation, and of a weaker field along the axis.

(1) INTRODUCTION

The history of the hollow cathode, so far as analytical approaches are concerned, goes back to 1951.[1,2] The concept of hollowness here refers to an almost perfectly enclosed cavity with a rather small exit. The present study was undertaken because of the need for an emitter which would give higher current densities and a smaller optical source than conventional oxide-coated cathodes. The latter could not take the severe centre loading when operated below narrow grid apertures.[3] On the other hand, for reasonably large beam currents in very fine spots, extremely small grid apertures become an electron-optical necessity. The substitution of a cavity orifice for the physical surface seemed to alleviate the danger of excessive damage to the emitter, and the extent to which these suppositions were justified will be discussed. There is also a need for interpreting the hollowness in the emerging beam. An interpretation of the observations is attempted since a good physical picture is believed to be of greater usefulness to applied research than oversimplified isolated facts which cannot be integrated into a consistent mathematical concept.

(2) EXTERNAL CHARACTERISTICS

(2.1) The Normalized Hollow Cathode and the Test Diode

A shape was established (Fig. 1) which resulted in a rigid but inexpensive construction. The cavity walls were coated

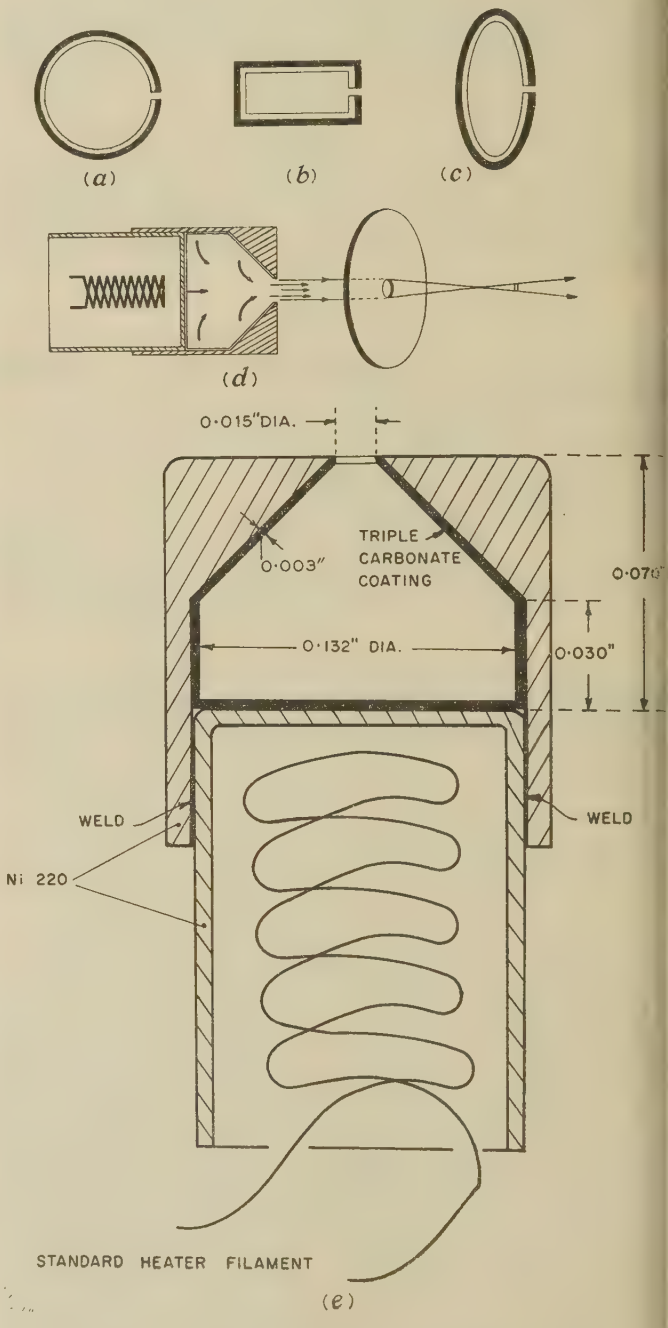


Fig. 1.—Internally coated cavities of different shape.

(*a*) Spherical.
(*b*) Cylindrical.
(*c*) Ellipsoid.
(*d*) With negative control grid.
(*e*) Normalized miniature hollow cathode.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Dr. Sandor is with the General Telephone and Electronics Laboratories, Inc., New York, U.S.A.

conventional barium-strontium-calcium carbonates which oken down and activated in the usual way. The control- perture is shown for half-tone writing guns. The total (inner) area is 0·273 cm². Inside the sleeve is mounted andard picture-tube heater, designed for 6·3 volts and np. An input power of 9 watts is required for a brightness rature of 850° C on the cap (920° C true temperature). orifice diameter was selected as the best compromise en emission current and a small virtual source.

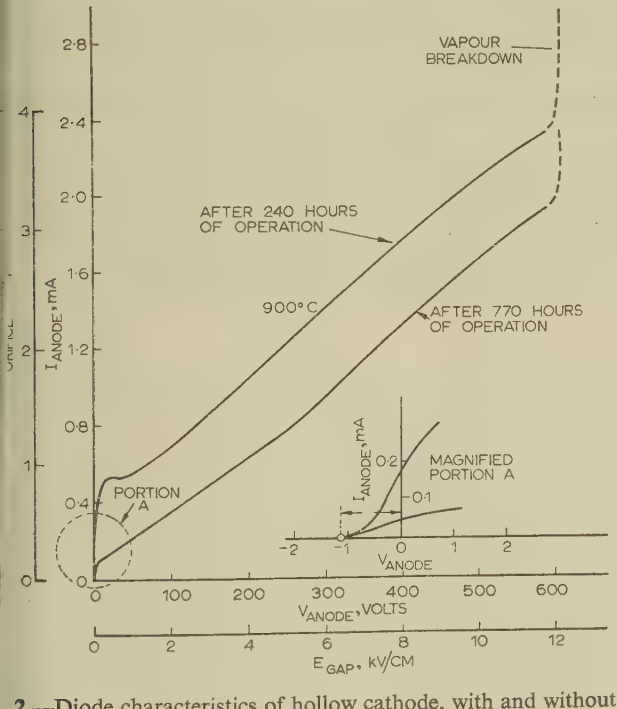
nstant mechanical parameters in the diode structure were ained by using the same electrode assembly in all tests, y employing a spring-loaded cathode-ceramic mount with et lock.

(2.2) Crystallization Effects

er activation a flaky, white crystalline formation of about in width develops along the orifice edges, as has been ved by Poole,[4] and this remains throughout life; this it seems to be the result of sublimation of oxide components g activation. In all cases a clear hexagonal crystallization n on nickel was obtained, spreading over the entire cap e. In calculating emission current densities at the orifice, he constricted 0·011 in-diameter hole was employed.

(2.3) Diode Characteristics

. 2 shows diode characteristics measured against an ired nickel anode plate with a 0·040 in diameter hole at a



2.—Diode characteristics of hollow cathode, with and without lip emission.

Temperature = 900° C brightness.
Cathode-anode spacing = 0·020 in.
Cavity side and bottom coated.

nce of 0·040 in. The cathode performed according to the r characteristic for 240 hours, then changed to the lower e. The initial characteristic shows the customary approxi- 3/2-power rise as seen in the enlarged initial portion; a saturation plateau is reached. From here on a more or linear increase of emission with anode voltage occurs until field strength of 12 kV/cm, measured between solid parallel faces, a sudden electrical gas discharge, caused by the high field at the orifice edge, breaks the trend. In the cavity an appreciable composite vapour pressure of the coating components can build up at an elevated temperature, while in the case of the planar conventional cathode immediate dissipation would follow.

Emerging from the orifice, a faint bluish beam becomes visible and remains equal to the orifice diameter; this constitutes an indirect method of showing that the electron rays are practically collimated, although some gas focusing might also contribute. During such vapour excitation the overall vacuum in the tube remains at about 10^{-6} mm Hg, judging by the condition of the getter mirror. The luminescent beam essentially coincides with the true electron paths. By collision and recombination processes the carriers of the main atomic mass of positive ions will establish a counter-flow and thus, with the anode voltage applied, the effective loss of vapour substance from within the cavity is compensated. The active coating, however, will suffer severe damage by positive-ion bombardment. This condition can be prevented by keeping the anode potential always below the firing point. The presence of energetic ions is manifested in an elongation of the round cathode orifice by a factor of 2. Positive ions are focused back by the slightly tilted anode aperture, so that they hit the top plate off-centre.

The final load figure, limited only by the vapour breakdown, was about 4 amp/cm² at a brightness temperature of 900° C (965° C true temperature), which is much higher than regular oxide cathodes can sustain during life.

(2.4) Characteristics Related to Time

From the shape of the initial (upper) curve in Fig. 2 one can conclude that the emitting source is at first close to the anode, i.e. to the orifice lip. With increasing voltage, an almost linear cavity emission becomes predominant in addition to the constant saturation current from the lip.

If the voltage required for saturated emission from the lip is exceeded, deactivation occurs.[5] This is particularly true if strong fields develop at the sublimated needle crystals, since emitting sites are left without negative space-charge protection and become vulnerable to ion bombardment. The current function then falls to run essentially parallel to the upper curve. The initial 3/2-power feature and the saturation plateau disappear, but the point of vapour ignition must remain the same, for the field and vapour pressure are unchanged.

(2.5) Temperature Dependence

Temperature-sensitive operation is shown in Fig. 3 for 850° and 900° C for a cathode-anode spacing of 0·002 in and an anode aperture of 0·015 in, with lip emission eliminated. The vapour condition starts now at a much higher field than with larger spacings—about 42 kV/cm compared with 10 kV/cm for a 0·040 in cathode-anode spacing. The maximum load of the cathode is accordingly extended from 5 to 10 amp/cm² for 900° C. This may have two causes: first, there is the very close conductive boundary (anode) at which charge annihilation by recombination and sink action would readily occur; secondly, the shortened cathode-anode distance is less than the mean free path of electrons required for collision ionization.

(2.6) Life Characteristics

Fig. 4 shows emission current for the preceding structure plotted against hours of continuous load. The load capacity improved by long ageing, so that about 7 amp/cm² could be constantly obtained for about 800 hours at 850° C brightness for longer life. From here on the emission falls gradually, but

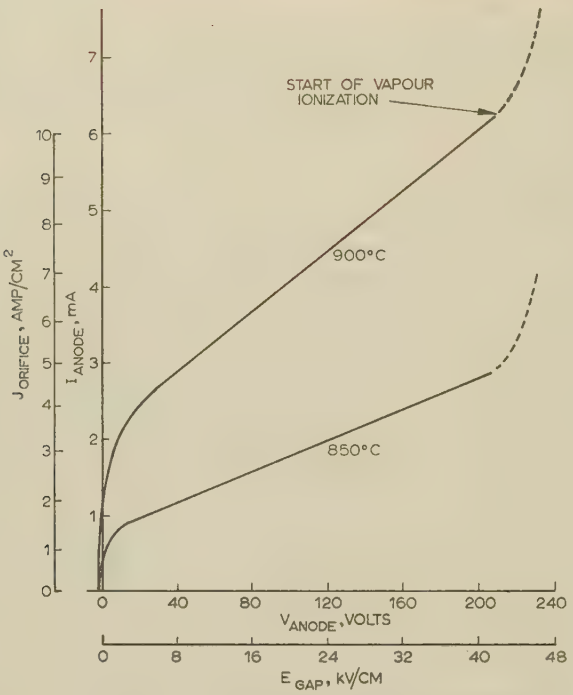


Fig. 3.—Temperature dependence of pure cavity emission.
Cathode-anode spacing = 0·002 in.
Anode aperture = 0·015 in diameter.

still maintains a value of 3 amp/cm^2 at the 1600-hour point. One factor very likely to cause the limited life span is a slow positive-ion damage.

(2.7) Modifying Cavity Functions

Fig. 5 shows different cavity modifications used in the tests. First, the standard shape with sharp feather lip, side wall coated (*a*) and then fully coated (*b*) is shown. These two configurations are practically equivalent and represent the best compromise so far as emission characteristics are concerned. Originall uncoated areas are automatically coated within the hot cavi by sublimation. Then, a hollow cap with four ventilation hol is depicted, (*c*), letting vapour diffuse rapidly into vacuum t prevent early ionization. The field strength can, indeed, b almost doubled before the gas discharge occurs, but emission not improved; on the contrary, it remains lower than in th former case. Moreover, an up-and-down cycling of emissio was observed, with a period of a few hundred hours. This apparently due to a slow spontaneous reactivation followed b pressure recovery which might produce more ion damage.

The next version, (*d*), was supposed to alleviate the critic vapour ignition by rounding the lip, thus lowering the fie Ignition was postponed, but longer channelling of the poten field into the cavity occurred requiring higher anode potentials

In the last version, (*e*), an attempt was made to increase i emission from two crossed nickel wires stretched over t orifice. Emission, solely from oxide sublimation on the wi was soon deactivated because of the voltage applied in exc of saturation, while the orifice was almost clogged by t sublimate. Field penetration into the cavity was sharply redu by the shielding wires and thus the vapour complex igni prematurely.

(2.8) Grid-Controlled Operation

The next step was to apply the hollow cathode to a nor television picture tube. Fig. 6 shows the current/volt characteristics of this arrangement. Cut-off voltages are t same as in the original tube, within the limits of usual toleran Emission currents are only of the order of microamperes inst of the milliamperes that can be obtained from planar cathod This lack of low-voltage response is basically caused by the stro potential shielding of the metallic cathode top and by the fa that the source of emission recedes into the cavity as increasi current is drawn. Accordingly, the transconductance is inheren lower with hollow cathodes; in the present case only 1·2 μA/vo are obtained as compared to 6·5 μA/volt with the planar oxi cathode.

Focusing was obtained on the screen, but strong spheri

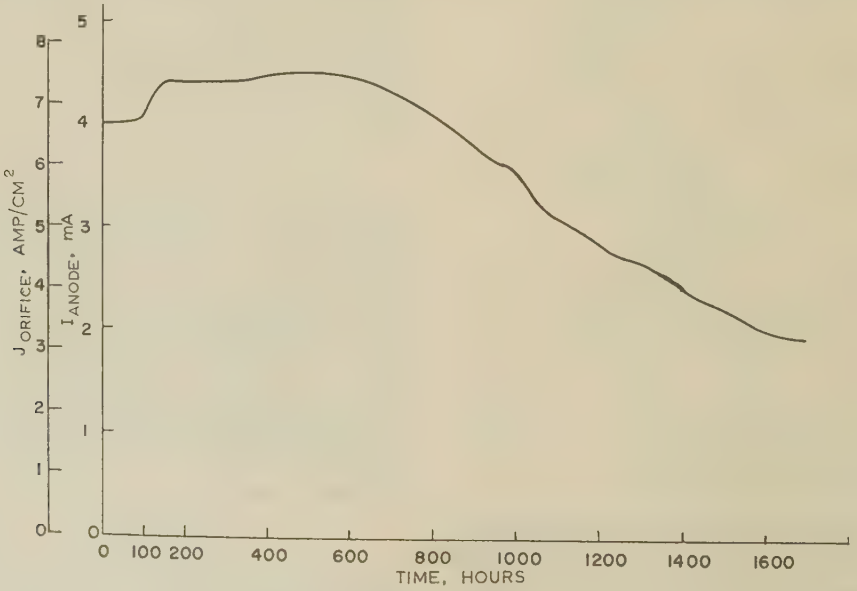


Fig. 4.—Life-test results of hollow cathode after optimum ageing.
Temperature = 850° C brightness.
V_{anode} = 210 volts.
E = 42 kV/cm in cathode-anode gap, measured between solid planes.
Cavity fully coated.

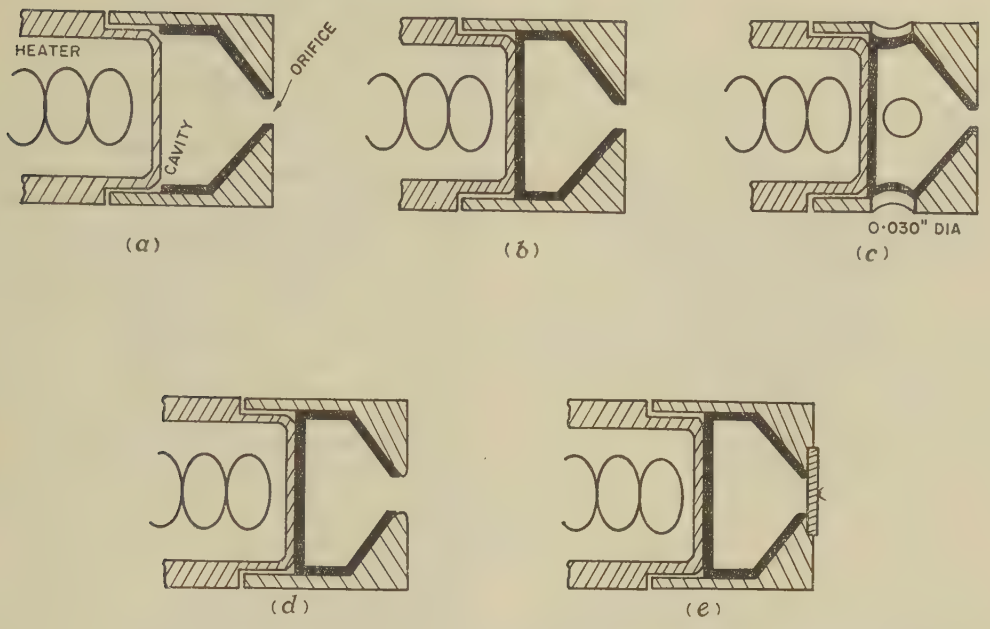


Fig. 5.—Variations of cavity and orifice concept.
(*a*) Side coated.
(*b*) Fully coated.
(*c*) With four side openings.
(*d*) Rounded lip.
(*e*) Cross wires over orifice.

astigmatic aberrations occurred in the spot, owing to the nation of a second (orifice) lens at the back of the grid.

MECHANISMS OF EMISSION AND BEAM FORMATION

(3.1) Internal Emission

irst, the problem of the emission mechanism within the cavity attacked, since nothing has been published so far except for ie speculations.[6] The main concern was whether there is a ic difference between small-surface high-field emission and e-surface low-field emission within a practically enclosed ce, and how the results compare with cavity emission to an rnal anode.

hollow cathode, as used previously, was equipped with a gsten wire of 0·010 in diameter in the centre of the cavity, ching through an insulating ceramic plug down the axis of heater coil. After proper activation and with the centre e negative with respect to the cathode cap, an emission current vs from the wire as oxide deposits form there from cavity ours. The total emitting area of wire and vapour shield is ut 0·01 cm^2. With the polarities reversed, the cavity walls it and the wire serves as a collector.

ig. 7 shows the measurements for different cavity tempera-es. When the wire was positive, emission curves were obtained ich qualitatively resemble those taken with the wire negative, in with the somewhat degenerate 3/2-power rise, but with aller total currents; saturation levels are apparently non-stent. It is believed that saturation would also show with walls as emitter if the voltage could be made high enough hout vapour ionization. The unevenness of the curve in the t quadrant is caused by progressive saturation along the ity wall. All of the curves show practically zero current, asured down to 10^{-8} amp, when the potential of the wire is 7 volt positive with respect to the cavity wall; this voltage is noticeably affected by changes in cavity temperature. One assume this effect to be the result of the non-symmetrical metry and differences between the temperatures and work ctions of the two electrodes. Although slow thermal ions

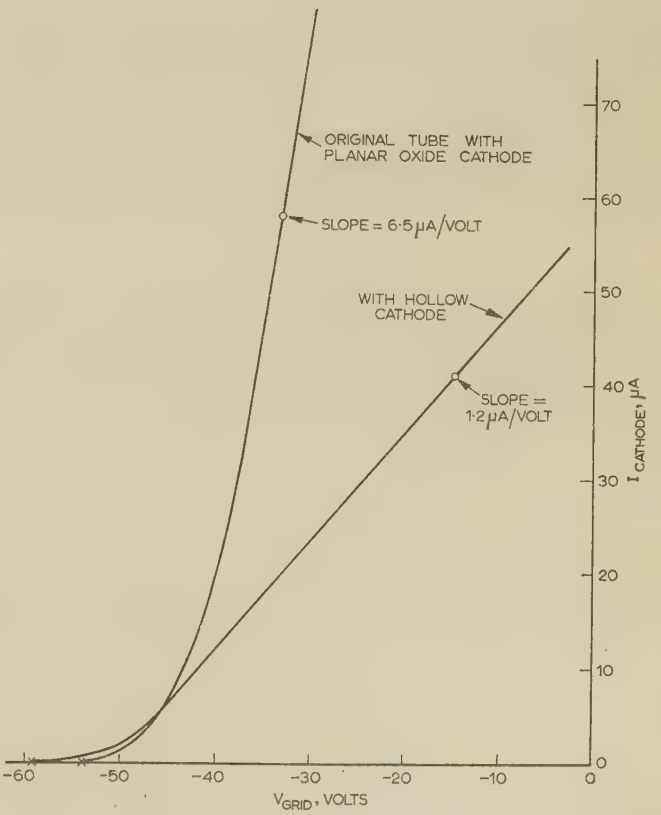


Fig. 6.—Comparison of grid-control characteristics of hollow cathode in picture tube.
Temperature = 800° C brightness.
Cathode-grid spacing = 0·005 5 in.
Grid aperture = 0·035 in diameter.
V_{anode} = 14 kV.
Second grid potential = + 300 volts.

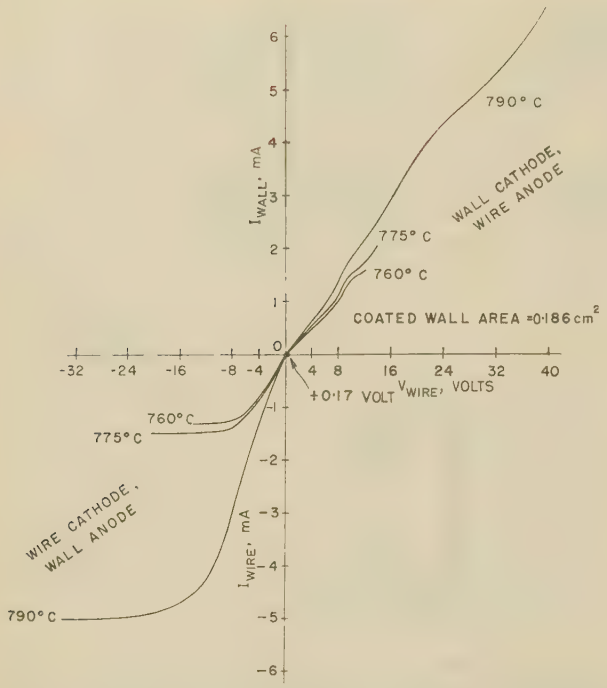


Fig. 7.—Internal emission characteristics of hollow cathode with centre wire.

are formed in the alkaline-earth/metal components of the oxide coating, their space-charge neutralizing effect is negligibly small as expressed by the shape of the undisturbed internal characteristics.

(3.2) Evidence for Modified External Mechanism

Since the beam cut-off voltages were practically identical with those measured with planar oxide cathodes, the location of the virtual planes of emission at quasi-zero current must be practically identical for hollow and planar cathodes. But with increasing field the current in the planar case increases much more sharply. The conclusion is seemingly obvious: the location of the virtual cathode remains nearly fixed for the planar cathode but recedes into the voluminous cavity in the hollow case.[6] As it recedes, it approaches the inner walls with an increasingly larger area. The shorter distances cause early current saturations as compared to small currents from distant areas. This proportion determines how large the total current at any given anode voltage will be.

In Fig. 8 a mechanism of emission is suggested on the basis of the former evidence. Since the highest electron density occurs at the orifice where no electron counter-pressure exists, the potential minimum will also develop there. If currents are drawn, the space-charge conditions are modified and the potential minimum recedes into the cavity. In its entirety, the virtual-cathode surface always represents a continuum of unequal potential minima because of the non-planar emitter geometry. Its local value and position will coincide with that of the wall itself where saturation proceeds. Ions will also strike the coating with corresponding energy and will cause cumulative damage.

Why the experiment with the perforated cavity wall [Fig. 5(b)] resulted in lower emission currents for the same external anode voltage now seems plausible. Since the cavity walls were punctured, the electron 'gas' lost in density (negative space charge) by electron leakage and the position of the potential minimum was confined to the vicinity of the inner wall surface. The anode field could not penetrate efficiently far enough to draw equal primary currents. Only the current maxima of the observed long-period emission cycles (see Section 2.7) are considered when referring to lowered current values as compared with the non-perforated-wall case.

(3.3) Dynamic Space-Charge Limitation

The surface V_{min} in Fig. 8 depicts qualitatively the location

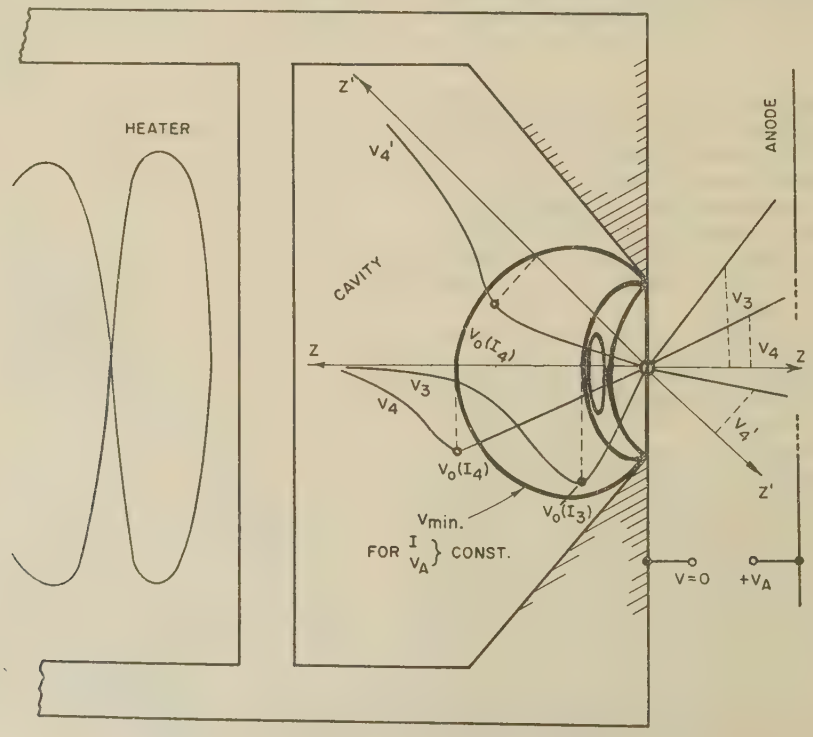


Fig. 8.—Pictorial model of virtual cathode formation.

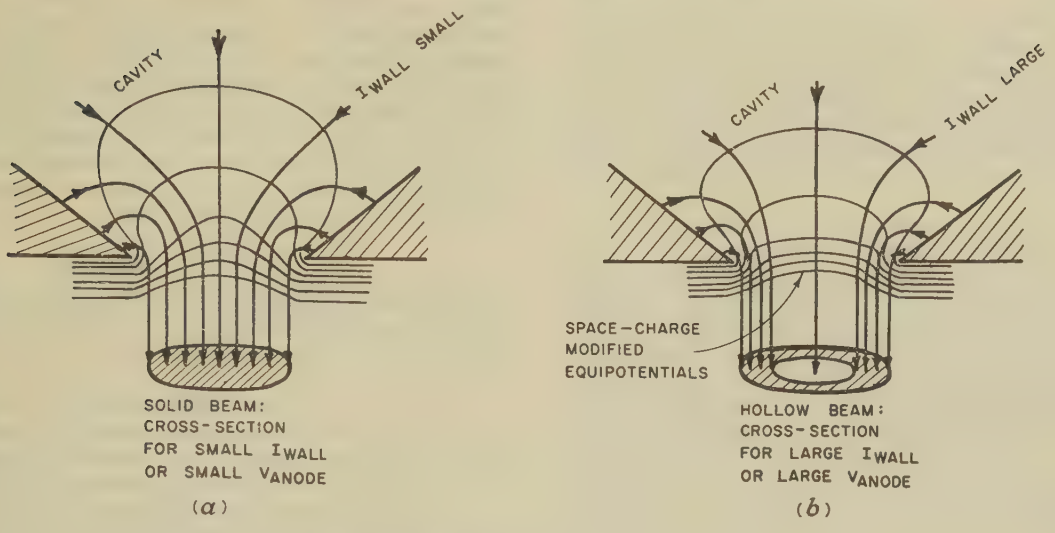


Fig. 9.—Model for hollow beam formation.
(*a*) Hollowness is absent.
(*b*) Hollowness formed by ray bunching at high current.

ne of the possible virtual-cathode planes comprising all ıtial minima, V_0, for a certain current, I. As the pene- ıg anode field, V_A, increases, both the emission current, nd its density at the orifice increase. Higher current drain rs the space-charge and shifts the virtual cathode backward the cavity. Its surface will expand roughly by a square tion, while the field, also advancing inward, will increase at ıch lower rate.

radually decreasing currents towards the cavity axis and asing saturation of wall areas will add to the complexity ıe emission mechanism. The resulting characteristics can pproximated by an initial 3/2-power rise, followed by a ıer 2/3-power function and finally by a lower, more or less r increase of the total anode current. In open planar odes practically no room is left for motion and expansion ıe potential minima because of the lack of an extensive e-charge-saturated electron-'gas' volume, the virtual-cathode ıce being almost identical with the thermionic surface itself. ly saturating areas are non-existent, and progressive weaken- of the field by apertures does not occur. Thus, with the ow cathode space-charge limitation of electron transmission mes absent by compensation rather than by elimination.

ıis trend is offset only in the very initial portion near = — 1·2 volts (see Fig. 2) as the virtual-cathode surface not expand within the cylindrical orifice of finite length. easonable physical picture of the emission mechanism, uding progressive or final saturation effects, could be ined by formulating a current/voltage relation for an valent diode with a complex dynamic perveance. However, ıified and explicit mathematical treatment does not appear ible.

(3.4) Mechanism of Beam Hollowness

bservations of the beam cross-section were made by direct ection of images on a screen. Fig. 9(*a*) shows the initial ıtion with small currents resulting in a solid beam cross- on.

ll the electrostatic force lines must necessarily pass through orifice, and thus the electrons will essentially follow their s because of the minute refractive power of the internal for the prevailing low electron velocities. The much ıger external field, on the other hand, makes the parabolic trajectories become nearly normal to the orifice plane. Hence, initial velocity differences, as well as initial angular spreading of the electrons, are also largely compensated.

If larger currents are drawn, the shape of the original equipotential surfaces is modified by increased space-charge [Fig. 9(*b*)]. As the equipotentials flatten out and move closer together, the refractive power becomes stronger, bunching the trajectories near the orifice edge. Another contributing factor is the receding virtual-cathode plane. Less current is drawn from its centre, while from the sides higher emission will result because of the shorter orifice-to-wall distances. Furthermore, lens aberrations and space-charge in the beam-processing elements will redistribute the original current densities. The general tendency will be to smear out the distinct ring shape.

(3.5) Spot-Size Comparison

Finally, Fig. 10 shows how beam spot sizes in a television picture tube compare. The spot size with the planar oxide cathode varies moderately with increasing beam current, while the hollow cathode produces rapidly increasing spots. Spherical aberration of the double-hole lens between grid aperture and

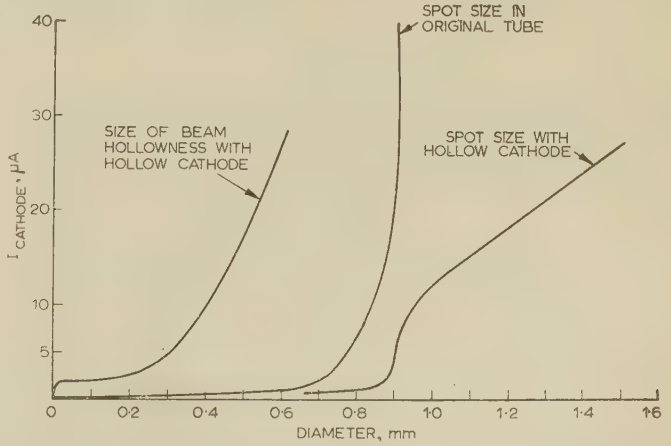


Fig. 10.—Spot size and beam hollowness with the hollow cathode in a picture-tube gun: comparison with planar-cathode performance.
The spot size is taken between opposite spot boundaries of 5% peak brightness.

cathode orifice produces this discrepancy, especially at higher current densities when an excessively large first-beam crossover is formed by space-charge repulsion. As already pointed out, it is not the angular spread of the electron paths that is responsible for the spot distortions, but mainly optical aberrations. Beam hollowness starts practically above 2 μA of total beam current. It increases slower than the spot size except in the beginning.

(4) PRACTICAL CONCLUSIONS

In conclusion, one would suggest the use of positive instead of negative grid control in order to enhance the total beam current by increasing the penetrating voltage, for a low intensity-control sensitivity appears to be inherent to the cavity principle. Where high loading and a narrow beam is the objective, open electrode configurations of the diode type or the Pierce-gun type are preferred in order to generate adequate potentials at the orifice. Sublimation of active material at the orifice edge occurs, resulting in large-diameter background images of ring shapes associated with fixed current components up to the point of deactivation.

High emission densities of about 7 amp/cm^2 can be utilized for long life at a brightness temperature of 850° C if the voltage region, where vapour ignition would start, is avoided. However, the life of coated hollow cathodes is necessarily shortened by a pronounced positive-ion activity as compared with open emitters.

Beam hollowness can become an important tool in microwave tube technology, where the core of a solid non-Brillouin type beam, consisting of low-velocity electrons, would be an undesirable feature. High-density ring-shaped narrow beam preferably constrained by an axial magnetic field, might be of great value in microwave amplifiers and oscillators with lo interaction regions.

Extremely small spot sizes, or very-small-diameter bea carrying heavy currents do not seem realizable in view of t basically high spherical aberration and astigmatism of t orifice lens.

(5) ACKNOWLEDGMENT

The author is indebted to Mr. L. D. Cohen, who was mo helpful in the experimental work.

(6) REFERENCES

(1) Holhouser, D. F.: Report of Electric Engineering Labo tory, University of Illinois, No. XIX-13, 13th July, 19
(2) Epsztein, C., Fried, C., and Smullin, L. D.: Prog Report of the Electronics Research Laboratory, Ma chusetts Institute of Technology, 15th October, 1952.
(3) Sandor, A.: 'Emission Decay Below Narrow Grid A tures', *Le Vide*, 1960, **15**, p. 373.
(4) Poole, K. M.: 'Emission from Hollow Cathodes', *Jou of Applied Physics*, 1955, **26**, p. 1176.
(5) Sandor, A.: 'Improvement on Oxide Cathode Stability Controlled Electrolysis', *Le Vide* (to be published).
(6) Brunn, K.: 'An Investigation of the Hollow Sphe Cathode', Dissertation, University of Illinois, June, 19

385.032.213.13

The Institution of Electrical Engineers
Paper No. 3363 E
Jan. 1961

ESSED-OXIDE NICKEL-MATRIX CATHODE BELOW APERTURED ELECTRODES

By A. SANDOR, M.E.E., D.E.Sc.

(*The paper was first received 12th October, 1959, and in revised form 24th March, 1960.*)

SUMMARY

rbonyl-nickel powder pressed and sintered with alkaline-earth e carbonates serves as an experimental emitter. The general viour of this dispenser-type cathode under the severe field and ocusing conditions below closely spaced and apertured electrodes vestigated.

ocally confined ionization processes in the alkaline vapour com- nts will develop. The ion-focusing ability of the aperture, par- arly if the surface-protecting space charge is removed, enhances oration. A spontaneous rise in cathode temperature accompanies effect, leading to more vapour generation.

eep pit formations will invalidate the advantage believed to exist well-determinable surface profile. Negative grid control produces ar ionization effects because of the large centre field below the ture.

ne self-sustaining discharge, starting just beyond the saturation l, will last as long as sufficient alkaline metal is available on the ice. Three discharge peaks were measured with gradually ased anode voltage, indicating the replenishing ability from within matrix. A faster recovery from the discharge state at higher peratures is explained by a higher deionization rate, owing to a e intense vaporization of nickel.

nder the severe aperturing conditions the cathode is inferior to e-coated cathodes and also to dispenser-type cathodes with built-in nical reaction and vapour controls.

(1) INTRODUCTION

he type of cathode under investigation is an experimental duct. No efforts were made to achieve any specific mionic properties; rather, emphasis was placed on studies ome aspects common to dispenser cathodes, particularly r behaviour in practical electron tubes with apertured elec- les. This work permitted qualitative conclusions to be drawn he design and functioning of similar cathodes.

he structure is based upon the dispenser principle, by which active substances such as barium, strontium and calcium he surface are continuously replenished from a voluminous rix. The skeleton consists of sintered nickel powder, retain- the oxides of the active metals in microscopic and mainly rconnected pores.[1] Incomplete activation and ageing, i.e. erfect freeing and arranging of active metal atoms in the oxide ce, gives the oxide bulk a chance to activate continuously er the influence of temperature after the tube is sealed off. omatic control of oxide reduction and evaporation rate is rent in this process. The freed active metals migrate slowly he cathode surface following the mechanism of diffusion ugh solid solutions, of surface migration and Knudsen flow ugh interconnected capillaries.

uch a structure should be capable of using high temperatures hence of delivering higher emission currents with a long life, tly because of its lessened resistance and partly because of the ss of active oxides stored and protected in the micro- leton of the matrix. Although the rate of evaporation will ease because of the higher temperatures employed, the voluminous replenishing mechanism, in conjunction with the pore-controlled rate of surface coverage and the protection which the emitting substance receives in deep capillaries, might still offer better life expectancy.

As implied in the foregoing, the higher temperature, to which emission is very sensitive, is the decisive factor in all dispenser cathodes and not the work function or the electron-concentration factor *per se*. These characteristics are even worse than with optimum oxide monolayers,[1] indicating that the atoms of the bare base metal, in close contact with active surface layers of alkaline-earth metals, will alter the favourable surface-dipole effect encountered in oxide bulks.

Nickel powder was chosen instead of refractory metals because of simpler fabrication methods. At low temperatures it can easily be sintered on to the cathode sleeve—an arrangement most favourable for heat conduction and mechanical stability. Machining is as easy as it is with solid nickel. Furthermore, in conjunction with the oxides it gives a work function which is lower than in any combination with refractory metal matrices.

The temperature advantage also constitutes the crucial factor in the question whether this type of cathode, loadwise and lifewise, is superior to the non-dispenser type when used with closely spaced and apertured electrodes.

Small openings were used in the diode plates close to the cathodes to simulate apertures of gun systems. Such tests are much more severe than with solid counter-electrodes or with open apertures. The narrow control-grid aperture of an electron gun harms the cathode centre through the effect of high fields and ion activity. The apertured anode of a diode causes similar damage, because the inverted equipotential surfaces focus ions back on to the cathode with the lowest field strength in the cathode centre; ionization is initiated by the high field at the aperture edges.

All measurements were taken under d.c. operation, although pulsed-grid conditions might produce more favourable characteristics and less cathode deterioration because of periodic fluctuations of the ion density and of the affected area.

(2) CATHODE STRUCTURE

As shown in Fig. 1, the cathode is shaped and dimensioned like the oxide-coated sleeves used in cathode-ray tube guns, with an active pellet of 0·134 in diameter and 0·020 in thickness fused to its top. Carbonyl-nickel powder is precompressed in a die

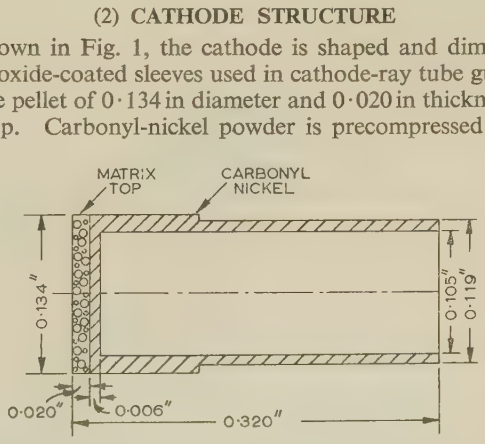


Fig. 1.—Cross-section of finished matrix cathode.

ritten contributions on papers published without being read at meetings are ed for consideration with a view to publication.

: Sandor is with the General Telephone and Electronics Laboratories, Inc., York, U.S.A.

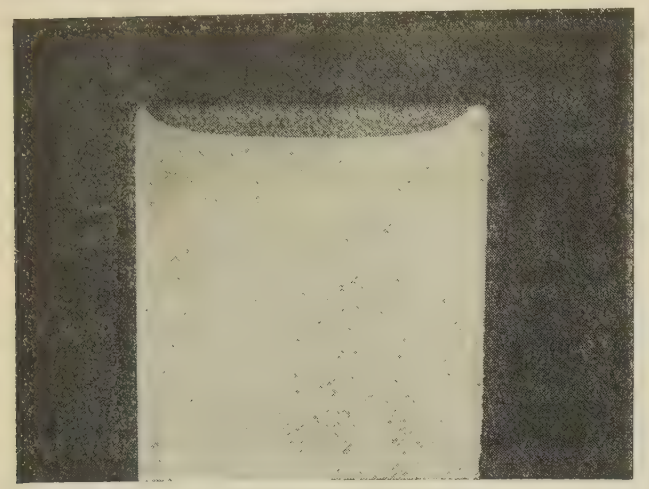


Fig. 2.—Solid nickel slug with pressed and sintered cathode pellet before machining.

Magnification, ×13.

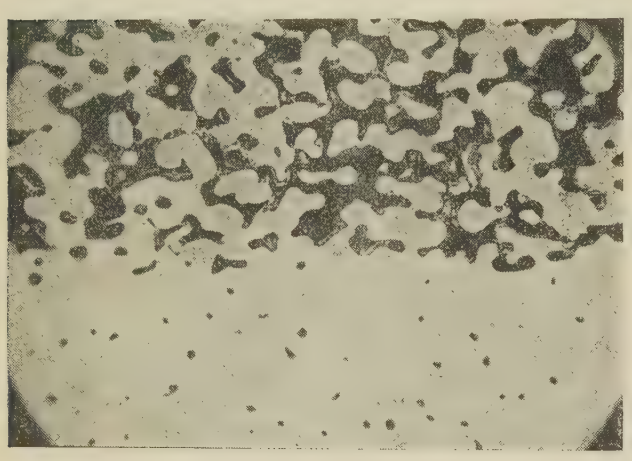


Fig. 3.—Micrograph of cathode pellet in cross-section showing interface region.

Dark areas represent pores filled with alkaline oxides; light areas are the nickel body.

Magnification factor, ×250.

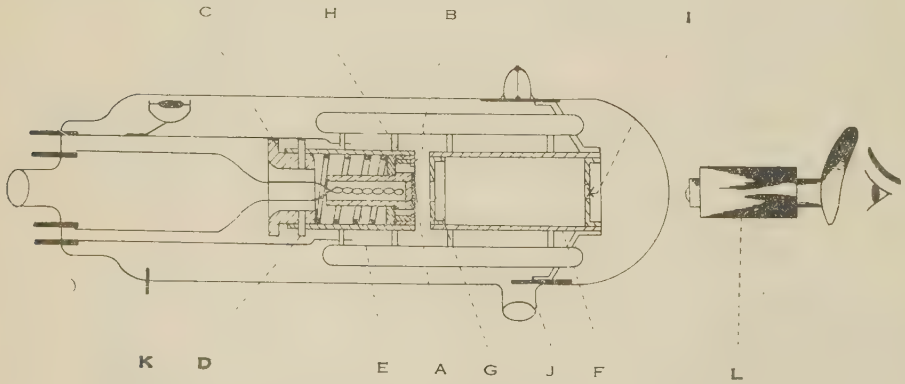


Fig. 4.—Semi-demountable apertured diode used for testing cathodes.

A. Matrix cathode.
B. Potential shield.
C. Loading plug.
D. Locking pins.
E. Loading spring.
F. Anode cylinder.
G. Aperture.
H. Spacer.
I. Tungsten cross-wires.
J. Outlet for enhanced nitrogen flow.
K. Line along which tube is cut and resealed.
L. Optical pyrometer.

to form a fairly dense slug and is then sintered in a protective argon atmosphere for ½ hour for final densification. Triple carbonates of barium, strontium and calcium are mixed with nickel powder in the ratio 1 : 3 by weight and poured on top of the slug in a die which encloses and extends above the slug. This top is now compressed with a pressure of 20 tons/in^2 and then sintered in argon at 1 300° C for ½ hour, when the carbonates release carbon dioxide and are converted into the respective oxides. A hollow cylinder is then machined from the composite nickel body to receive the heater.

Fig. 2 shows an enlarged cross-section of the slug, and Fig. 3 a micrograph around the interface between the cathode pellet and the nickel body. The dark areas are the porous portions of the pellet filled by the oxides. As is evident, the bond between pellet and nickel body is very strong.

In order to protect the slug from atmospheric influence, especially moisture, the pellet portion is dipped into a collodion solution to make the surface airtight. After machining, the finished cathode is stored in a desiccator. However, the oxides are apparently better protected by the dense micro-skeleton of nickel than are oxides directly exposed to air, which tend to convert quickly into quite stable hydroxides, accompanied by swelling.

The pellet thickness and the mixing proportion were arbitrarily chosen, as was the method of fabrication. When judging the performance of this design, these facts should be kept in mind, since there is room for improvement.

(3) THE TEST DIODE

A normalized diode system with an anode aperture of 0·040 in diameter was provided for all investigations, in order to ensure consistency of important parameters. After the tube has been cut open, its cathode can be exchanged and the tube resealed.

Fig. 4 is a schematic of the test diode as set up for optical pyrometry, the last aperture being provided with tungsten cross-wires, I, to act as electron catcher. The construction material is non-magnetic stainless steel.

The processing schedule is purposely extended in order to achieve thorough outgassing of the cathode and the electrodes under vacuum and to provide a better control of each phase. A pause at 1 100° C gives the residual locked-in gases a chance to dissipate thoroughly. The cathode is aged for 10 hours, but shorter processing schedules could certainly be determined.

The cathode/anode spacing is 0·040 in. All cathode temperatures quoted are brightness temperatures, the true temperature of the complex pellet structure being about 67–110° C higher, at least in the range (800–1 100° C) considered here. This is based on a radiation factor of about 30%.

(4) MEASUREMENTS

Two cathodes were investigated, one for continuous life and the other for behaviour under varying cumulative conditions. In this way certain anomalies could be shown, as well as their effect on each other. It appears, however, that preceding steps do not greatly affect the following ones with respect to final emission levels. This is a consequence of the slow but efficient

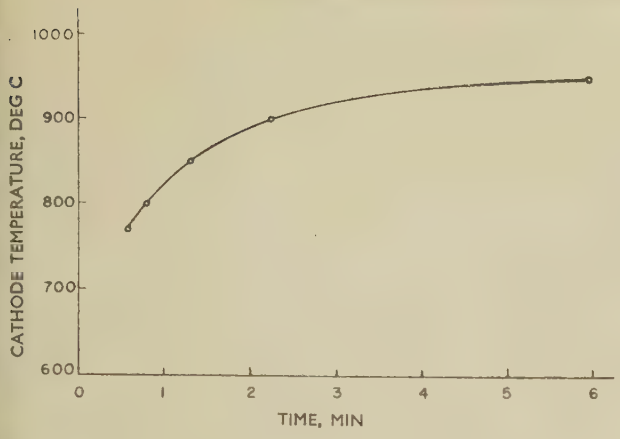


Fig. 5.—Heating characteristic of cathode for a constant filament input of 9·2 watts.

ecovery mechanism which backs up practically all the prelimi- naries, even severe maltreatments.

The effective cathode area used for computing the current oading was 0·091 cm^2, and uniform emission density was ssumed.

(4.1) Rise of Cathode Temperature with Time

The surface temperature of the cathode rises slowly, as shown n Fig. 5, because of the massive cathode body. The input of ·2 watts produces a temperature of 900° C after 2½ min, and he operating temperature of 950° C after 6 min, from the cold tart.

(4.2) Saturation Regions

As Fig. 6 shows, there are typical saturation knees, lying ower and shifting toward higher anode voltages as the cathode emperature is decreased. Since life at 1 050° C is significantly shortened, concentration on 950° C seems indicated for longer life. In this optimum case the saturation bend occurs at a field strength of approximately 1 673 volts/cm, measured between planar electrode faces, corresponding to an anode voltage of 170 volts and producing an average cathode load of 782 mA/cm^2 or 71 mA total current.

For any temperature, and when the anode voltage rises above the first onset of saturation, a sharp current peak develops which cannot be explained by simple thermionic concepts. With increasing anode voltage three consecutive peaks are observed with the 900° C curve, spaced 20 and 40 volts apart, respectively. The function connecting all peak onsets is shown as a broken curve, and the peaks are accompanied by a spontaneous increase in cathode surface temperature, in one case from 1 050 to 1 210° C. This is due to the impact of high-density positive ions and to resistive heating in the oxide plugs.

Local vapour concentration of the readily available active alkaline-earth metals prepares the ground for close-range ionization processes. Hence, after losing its protective space-charge above the saturation voltage, this type of cathode is apt to develop localized gas discharges of the self-sustaining type starting at the anode aperture edges of highest field gradient. The dense positive-ion beam will also compensate for the remainder of the negative space-charge in microscopic depressions.

The impact of the prefocused positive ions initiates a much heavier vapour phase than would occur with a solid anode. The active metals, with an inherently high vapour pressure of 5–20 mm Hg and a moderate ionization potential of 5 eV, will become temporarily exhausted and the discharge will cease. Although resonance ionization might be a possibility, nickel would generally produce an unfavourable ionization condition with a low partial local pressure of 10^{-6} mm Hg, a long mean free path and a higher ionization energy of 7·5 eV. The resultant pressure close to the cathode surface would actually be determined by the rate of evaporation and of the thermodynamically efficient atomic sputtering in dynamic equilibrium with the volumetric expansion in time and the sorption or condensation rate on parts of the tube. Neutral nickel atoms would actually help deionization, but with thermal evaporation supporting the violent sputter phases, dense nickel jets would diffuse into the vacuum and settle on cooler obstacles (electrical leakage). Thus a heavy loss of matrix substance will result.

After exhaustion of the active surface layers the discharge stops and the cathode temperature returns slowly to its nominal value. It takes 1 min at 1 050° C to return to the equilibrium level and 38 min at 850° C; initially it took only seconds to rise to the peaks. The faster recovery at higher temperatures is apparently due to a higher deionization rate by more numerous nickel atoms. The repeated recovery proves the existence of an efficient dispenser mechanism.

A quick recovery, as apparent in Fig. 6, does not always hold, because the same anode potential was maintained only until the emission fell to the previous level, the voltage then increasing again immediately; only in

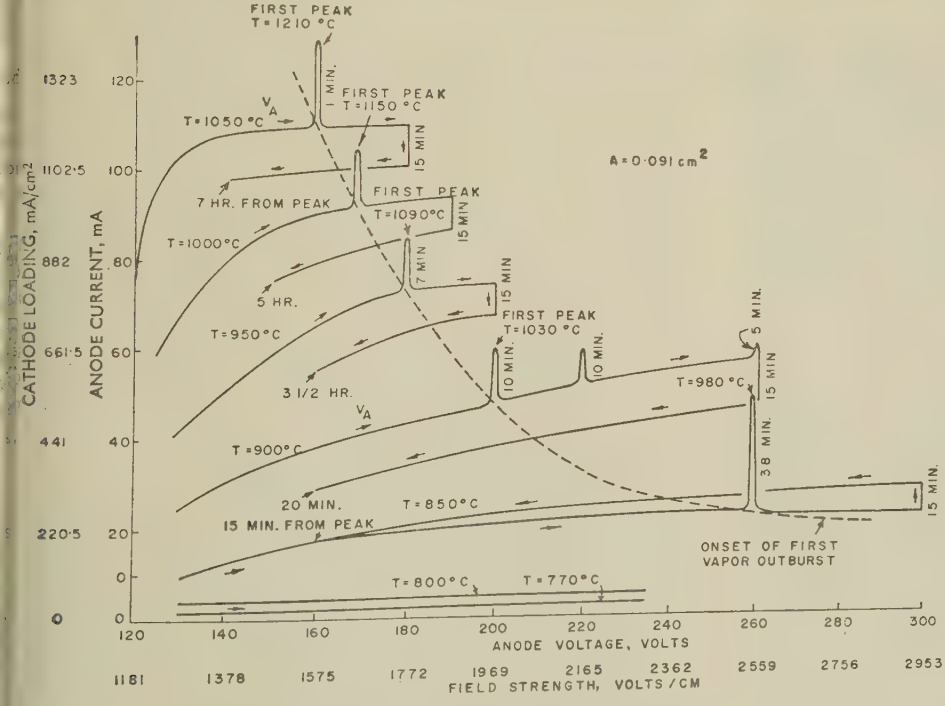


Fig. 6.—Saturation characteristics for different temperatures, showing the mechanism of vapour outbursts below an apertured anode.

Data taken after approximately 450 hours of tube operation at 950° C and an anode voltage of 160 volts.

the 900° C case, where the emission dropped below its previous level after the third outburst, was the anode potential kept constant longer. It seems that the compensation for the falling tendency is caused by the transport of active metals to the surface with increased anode voltage. If, however, the anode voltage is not quickly readjusted, the current continues to fall until exhaustion of the alkaline surface covering puts a necessary end to the vapour discharge. The drop of 8 mA, on the average, remained essentially constant, an exception being the 850° C characteristic, which improved temporarily.

The long-range fall of current is also associated with a deep pit formation in the pellet during vapour outbursts. Since the emission current decreases faster than inversely with the electrode distance, a sensitive decrease of current will follow, especially here, where the initial separation was only 0·040 in and the depth of the final destruction was about 0·010 in. The base of this dish-like pit, as shown in Fig. 7, is equal to the anode

Fig. 7.—Pit formation in the nickel matrix by positive alkaline-metal ions focused back on to the cathode surface by the anode lens.

Magnification, ×50.

aperture (0·040 in diameter). Its profile seems to be a replica of the focused ion stream, since the depth of sputtering will be approximately proportional to the ion current density causing it.

When the multiple outbursts at 900° C are given a chance to develop, deeper matrix regions are apparently affected. The decrease of peak heights seems to indicate such depth action yielding less and less active metals for evaporation before the main replenishing mechanism can become fully effective again. Those critical regions can be of no help in any gun operation, and thus closely spaced grids and apertures constitute risky elements.

(4.3) Optimum Emission Characteristics

The fall in emission after each vapour outburst does not affect the current permanently. Further improvement can be expected by approaching the ideal patchless monolayer condition, which occurs when the cathode is operated at 950° C and an anode voltage of 160 volts. This condition establishes the balance between the rate of replenishment, the rate of thermal surface evaporation and some sputter losses by a weak ionization process close to the surface. Realization seems to be accomplished when observing the ordinate values of the characteristics shown in Fig. 8, taken after the saturation curves in Fig. 6 were measured. All currents are higher for the same anode voltages, but the reproducibility of these characteristics was perfect. Maximum stable loadings of 1·2 amp/cm² at 1 000° C and 780 mA/cm² at 950° C were established as optima just before a full vapour discharge develops.

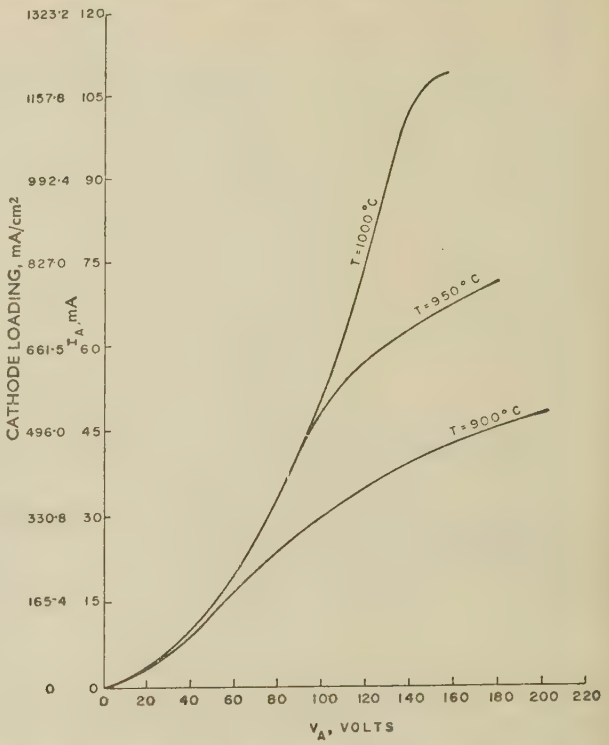


Fig. 8.—Current/voltage characteristics for different temperatures below the saturation bends.

Data taken after approximately 560 hours of operation at 950° C and an anode voltage of 160 volts.

(4.4) Life Characteristics

For the all-test cathode the life test was briefly interrupted for checking other phenomena (Figs. 6 and 8). The efficient reservoir property backed up all these violent excursions in due course, and another of the same cathode batch showed the actual differences when the auxiliary tests were omitted. The investigation on life for the all-test cathode was divided into two parts: one at 950° C and 160 volts anode voltage and the second at 1 050° C and 155 volts. The cathode for the continuous life test was operated at 950° C and 160 volts throughout. The life characteristics are shown in Fig. 9.

After ageing, which was accidentally started high in the region of vapour outbursts at 200 volts (first disturbance), the emission first fell briefly from about 660 to a minimum of 260 mA/cm², but recovered steadily to 665 mA/cm² by virtue of the dispenser property. This was followed by a slight decay corresponding to the second vapour outbursts. The total drop was now much smaller and the recovery much faster, because of the moderate anode voltage involved. This generally seems to hold.

After stabilization was obtained again, the temperature was abruptly raised to 1 050° C, the anode voltage being simultaneously lowered to 155 volts in order to circumvent possible vapour outbursts. After 160 hours of stable operation at a load of 1 190 mA/cm², a spontaneous vapour outburst put an end to the upward trend (fourth disturbance). A rapid increase of current (discharge) by 32 mA, followed by a sharp drop of 56 mA below the formerly stable level, put the characteristic back on a lower level. From here on the emission fell continually, with only a weak attempt to restabilize after 850 hours. Much later, another spontaneous vapour phase started (fifth disturbance), but the current returned to its base value after 60 hours of anomaly. This point marked the final exhaustion of the dispenser mechanism.

The width of the vapour time-base was then much broader

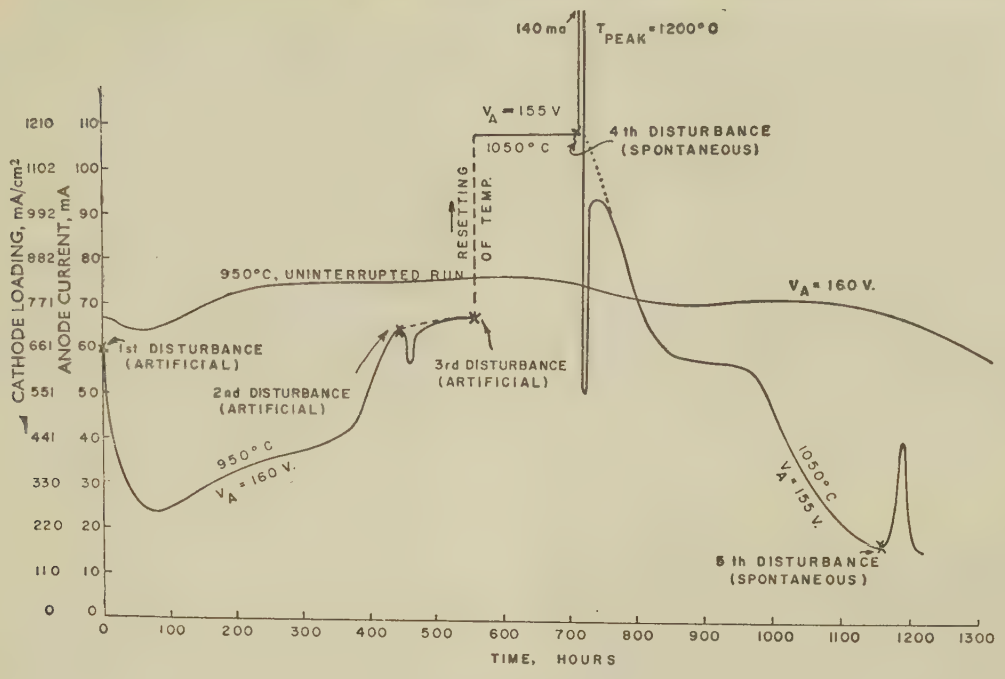


Fig. 9.—Curve showing ability of cathode to recover three times after heavy abuse by periodical outbursts of alkaline and nickel vapours.

Uninterrupted life run over 1 300 hours is shown in the upper curve.

than before (60 hours instead of minutes). No forceful ion activity in the thin vapour can build up, and the phase must expand in time. After the dispenser property approached exhaustion, the saturation level moved farther down and the bend of the characteristic moved to the left. Hence, the fourth and fifth disturbance were actually caused by the same mechanism of a continuously-falling saturation level arising from a loss of space-charge backing (active metal).

The compromise temperature of 950° C also led to a tolerable leakage rate by nickel sublimation when a vapour shield was pulled over the cathode sleeve to protect the ceramic disc. Furthermore, the warming-up time and the filament input assumed reasonable values, as can be seen from Fig. 5. However, the onset of pit formation, apparently due to a constantly existing still discharge in the narrow gap and enhanced by the ion-focusing property of the apertured anode, was still present. Measurements made between planar electrodes have not revealed the restricting factors, because of the lack of ion-focusing back into the cathode centre.

As is evident from Fig. 9, the life is acceptable when a more serious emission decay does not start before 1 200 hours. Intermittent operation or operation with a lower average direct current might result in a considerable extension of life.

(5) CONCLUDING DISCUSSION

The improvement of loadability by the use of higher temperatures encounters restrictions when narrow electrode apertures are employed, the most severe being that of excessive vapour generation in the alkaline-metal components above critical anode voltages, which leads to ionization states. Even reaction-controlled pressed-oxide dispenser cathodes of the tungsten-molybdenum type generate, in the much less severe case of solid planar anodes, a composite vapour complex which is still 20 times denser at their recommended temperature than that of the conventional oxide cathode at optimum long-life temperatures.[2] The sublimation rate is of the order of 4 instead of 0·21 $\mu g/cm^2/h$.

In addition, the proportion of the pure metal vapour (barium, strontium, calcium) to the corresponding oxide vapour is markedly shifted towards the metals. Since their partial pressure is much higher than that of the oxides and their ionization potential much lower, the vapours will readily ionize in the confined space. Heavy positive-ion activity on the cathode surface will result.

The rate of vapour generation in this pressed nickel-matrix cathode is even greater than in the refractory types because of the lack of chemical reduction controls. Its optimum operating temperature of 950° C, which was determined for apertured conditions, is limited by early ionization states. The temperature factor, otherwise so beneficial for the increase of thermionic emission, cannot be used efficiently, and the maximum useful yield is lower than that of well-prepared conventional oxide cathodes. To maintain a well-defined surface and constant spacing seems impossible even below the critical voltage point, because of the slow but steady surface damage by still discharges under an apertured and closely spaced electrode. The low-melting-point nickel will yield to vapour generation and pit formation more easily than matrix bodies based on refractory metals.

Applications in gun systems with negative-potential small grid apertures will be critical, since the cathode is incapable of producing sufficiently high loads, as required at the cathode centre, to maintain surface protection by space-charge.[3] The vapour condition will thus destroy the centre region. Its high rate of evaporation would also cause heavy sublimation on the grid, followed by excessive grid emission. It is important to consider the severity of testing conditions with apertured anodes in comparison with those in planar diode arrangements.

Since higher operating temperatures would increase the thermal noise, preference must be given to the oxide-coated cathode. With pulsed beams a gain was presumed, based on the fact that the overall bulk resistance (metal matrix) is reduced in comparison with the conventional oxide coating. However, this gain turns out to be of little final value, since the current

must still pass through the high-resistance donor depletion layer of only about 1 μ of the uppermost surface deposits. Consequently, current decay during a pulse of medium duration must and does occur, as it does with all dispenser type cathodes.[4,5] Low-frequency flicker noise would also be present.

Thus, dispenser cathodes in general seem to lose the technical interest which they initially attracted, so far as their application in closely spaced and narrow-apertured guns for high beam currents is concerned.

(6) ACKNOWLEDGMENTS

The author wishes to thank Dr. H. Bender, formerly of the Sylvania Metallurgy Research Laboratories, for supplying the cathode slug, and Mr. L. D. Cohen for his valuable assistance in preparing and performing the measurements.

(7) REFERENCES

(1) Beck, A. H. W.: 'High-Current-Density Thermionic Emitters: A Survey', *Proceedings I.E.E.*, Paper No. 2750 R, November 1958 (**106** B, p. 372).

(2) Coppola, P. P., and Hughes, R. C.: 'A New Pressed Dispenser Cathode', *Proceedings of the Institute of Radio Engineers*, 1956, **44**, p. 351.

(3) Sandor, A.: 'Emission Decay below Narrow Grid Apertures', *Le Vide*, 1960, **15**, p. 373.

(4) Nergaard, L. S.: 'Trends in Thinking about Thermionic Emitters', R.C.A. Industry Service Laboratory Reports, No. RB-56, June, 1956.

(5) Déjardin, G., Mesnard, G., and Uzan, R.: 'Pulsed Thermionic Emission from Thoria Cathodes', *Cahiers de Physique*, 1956, **10**, p. 1.

621.385.12.032.2

The Institution of Electrical Engineers
Paper No. 3344 E
Jan. 1961

ELECTRON EMISSION FROM COLD MAGNESIUM OXIDE

By H. N. DAGLISH, B.Sc., Ph.D., A.Inst.P., Graduate.

(*The paper was first received 17th May, and in revised form 2nd August*, 1960.)

SUMMARY

A porous layer of magnesium oxide may be used as a cold source of electrons in a vacuum tube, the emission depending upon the maintenance of a positive electric field through the oxide layer. Methods of making and processing cathodes of this type are described, and it is shown that stable emission is obtained only from cathodes in a fully oxidized state. The behaviour of these cathodes under various experimental conditions has been examined with particular reference to voltage and temperature effects.

(1) INTRODUCTION

A cold cathode coated with a thin layer of magnesium oxide may, in suitable circumstances, be used as a source of electrons in a vacuum tube. The emission phenomenon itself can be simply described. An electric field is established between the magnesium-oxide-coated cathode and an adjacent electrode, this electrode being positive with respect to the cathode. Some suitable external means is used to cause the loss of electrons from the cathode surface, and so cause the oxide surface potential to approach that of the positive electrode. If this potential is sufficiently large, the resulting field through the magnesium-oxide layer induces a small primary current which is unstable and short-lived. However, if the structure of the oxide layer is suitable, a much larger secondary current is produced by an avalanche mechanism, and since a high proportion of these secondary electrons is lost to the adjacent positive electrode, the surface of the oxide retains its positive potential. In these conditions the emission processes are self-sustaining.

The phenomenon is in many ways similar to the Malter effect. Malter[1] and Koller and Johnson[2] obtained field-enhanced secondary emission and short-lived self-sustained emission from very thin (10^{-3}–10^{-4} mm) layers of oxide (produced by oxidizing an aluminium surface). Nelson[3,4] showed that field-enhanced secondary emission drawn from an evaporated film of magnesium oxide was accompanied by the development of a positive potential at the surface of the oxide. Somewhat thicker emitting coatings of magnesium oxide were described by Jacobs and Dobischek and their co-workers[5–8] following their work on the high secondary-emission coefficient of the oxide.[9,10] Similar cathodes have also been described by Veenemans and van Zanten.[11] In most of these earlier experiments the secondary-emission coefficient of the magnesium oxide was further increased by treating the coating with caesium oxide. More recent work on practical cathodes has been described by Skellett and others.[12] The present paper describes some of the experimental work on cold emission from magnesium oxide carried out in the laboratories of the Post Office Research Station.

(2) EXPERIMENTAL ARRANGEMENTS

(2.1) Description of Valves

The basic experimental assembly was made from standard components as used in the current Post Office type-10P repeater valves.[13] It consisted of a rectangular box-type cathode core with an insulated internal heater, together with one or more helical grids. The heater was normally used only during processing of the cathode, and apart from experimental purposes, was not energized after the valve had been removed from the pumping system. The various electrodes were supported between the usual mica insulators and mounted on a glass pinch. This was sealed into a glass envelope for evacuation and processing. Soft lead glass was used throughout.

A typical diode assembly is shown in Fig. 1. It will be seen

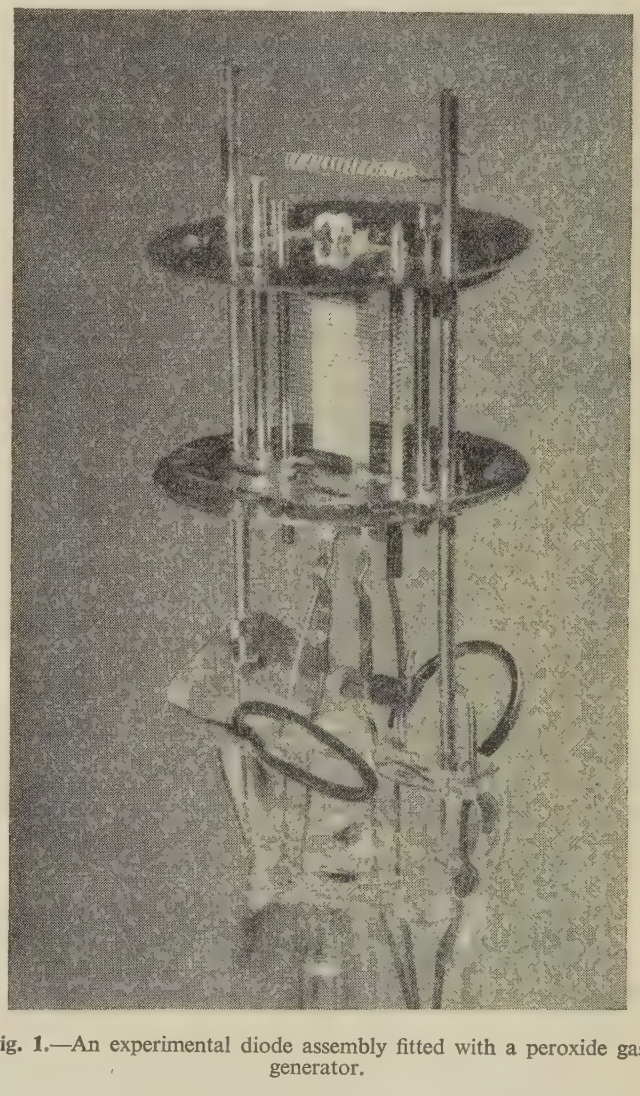


Fig. 1.—An experimental diode assembly fitted with a peroxide gas generator.

in this photograph that the support micas are fitted with metal shields. These are normally at cathode potential, and they have been found to discourage the accumulation of charge on the insulating surfaces of mica and glass. In their absence, the behaviour of the device becomes very dependent upon the

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Dr. Daglish is at the Post Office Research Station.

potential of the glass and upon the presence of objects outside the envelope altogether. As a precaution against such interference, many valves were built as triodes, with an outer grid electrode at cathode potential.

(2.2) Cathode Core Material

Most experimental cathodes were made from high-purity nickel, although successful cathodes were also made from 4% tungsten-nickel, and 'active' 0-type nickel. The cathodes were degreased with solvents, and stoved in hydrogen at 925° C. Pure platinum cores were also used, processed either in air or hydrogen.

(2.3) Composition and Structure of the Coating

The earliest experiments were carried out with a very simple coating technique. The metal cathodes were held in the white smoke produced by burning magnesium ribbon in air. This method is difficult to control, the thickness of the coatings varying considerably, while the oxide is very soft and easily damaged. Further work was therefore restricted to a more conventional spraying technique.

Various experimental spray pastes were used, each containing some of the following magnesium compounds: oxide, formate, oxalate and carbonate, together with an organic binder. Following a suggestion by Skellett, pastes were also used based on magnesium peroxide and carbonate in amyl acetate. Details of two typical pastes are included in Section 13.1.

The cathode cores were mounted in a suitable masking frame and sprayed with successive layers of cathode paste, the coating being dried with warm air between layers. The thickness of the coating was judged by eye, and the most satisfactory coatings were just thick enough to obscure all visible core metal. Such coatings weighed about 2–3 mg/cm.^{2} During the processing of the completed valves the various magnesium compounds were decomposed thermally to give coatings consisting entirely of magnesium oxide. About one third of the original weight was lost during this processing, the final coating being about 60–80 μ thick with a fluffy porous structure. An enlarged photograph of part of such a coating is shown in Fig. 2.

Provided that the processed coating has this type of porous structure, the exact composition of the spray paste is not critical, successful cathodes having been made with a variety of different spray formulations.

(2.4) Processing

The first experimental coatings were processed in a manner similar to that used for conventional barium–strontium-oxide cathodes. After evacuation and baking of the tubes for an hour, the cathodes were heated to decompose the magnesium compounds. Cathodes processed in this manner were found to become discoloured, the surface being apparently contaminated with carbon. No emission was obtained.

Various experiments were tried to prevent this contamination. It was found that if the tubes were evacuated only with a rotary pump during decomposition, the pressure was then usually sufficient to permit oxidation of the carbonaceous contamination. Cathodes processed in this way are capable of self-sustained emission, although not very satisfactorily.

Even greater improvement was obtained by the introduction of more vigorous oxidation during processing, the cathode being heated in various pressures of oxygen or air, after the conversion of the magnesium compounds to oxide. A typical process schedule is given in Section 13.2. The cathode temperatures quoted there are approximate only, because in practice the cathode temperatures are specified in terms of heater voltage.

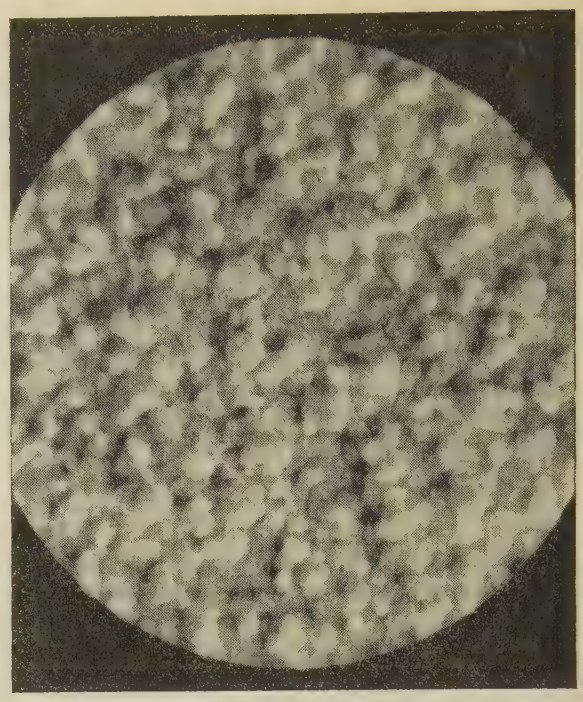


Fig. 2.—Part of a satisfactory coating (linear magnification ×50).

The purpose of the oxidation is twofold. All residual traces of carbon or organic material are converted to gases and removed. The oxidation also ensures that the magnesium oxide is in a fully oxidized condition, any variation from stoichiometric composition being towards excess oxygen rather than towards excess magnesium.

Many successful cathodes were produced to this scheduled treatment, irrespective of minor processing variations, such as reducing the pressure during the conversion of the magnesium compounds to oxide. However, combining this conversion with the high-pressure oxidation did impair the stability of the final emission. It was found easier to control the oxidation when using air rather than pure oxygen. Use of too high pressure or temperature resulted in premature failure of the tungsten heater. If the final outgassing was omitted, it was often found that the self-sustained emission was not very stable. The emission was not obtained continuously from the whole of the coating, but the active zone continuously flickered between several small areas.

(2.5) Starting the Emission

Emission of electrons from the cold magnesium oxide depends upon the establishment and maintenance of a positive potential on the oxide surface, and upon the existence of a suitable field outside the coating to remove the emitted electrons. This field is provided by applying a positive potential to the grid of the simple diode. No current passes when this potential is applied until primary emission is initiated.

There are several convenient ways in which this can be done. The simplest consists of irradiating the oxide surface with ultra-violet light, to expel photo-electrons from the oxide and so leave a positive potential on the surface. The method is reliable, but slow, since a considerable time may elapse before an adequate surface charge has accumulated. The other method which has been used is to allow a high-frequency discharge from a Tesla coil to play over the glass envelope. Emission usually commences immediately.

Other methods, such as flashing a tungsten filament near the cathode, or spraying electrons from a subsidiary source, require

ditional electrodes inside the envelope. These methods have erefore not been used much in the present experiments, hough they might be the most convenient for devices making e of the magnesium-oxide cathode.

A series ballast resistor is included in the electrical circuit to ibilize the emission, 10–50 kilohms being a suitable value. ie simple circuit is shown in Fig. 3.

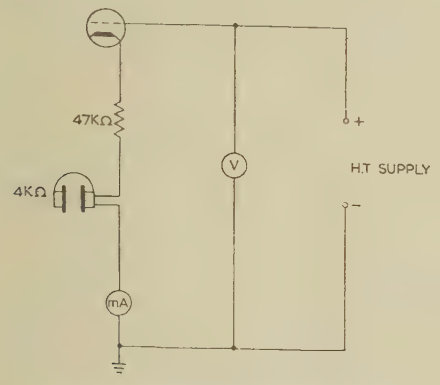


Fig. 3.—Simple experimental circuit.

(3) CATHODE PERFORMANCE

(3.1) General

The most noticeable feature of the emitting cathodes is the ight blue luminescence. (The band spectrum of this lumi-scence has been studied by Woods and Wright.[14]) With able cathodes whose performance is sufficiently good to be of actical value, examination under the microscope shows the hole surface of the coating to be covered by a relatively stable d uniform blue glow. Although the cathodes made by the rlier techniques behave in a less satisfactory manner, they vertheless reveal some interesting phenomena. It is worth scribing these because of the light they throw on the emission echanisms. These cathodes are described below as 'noisy' thodes.

(3.2) Noisy Cathodes

'Noisy' cathodes in general had coatings which were too thin bout 20–30 μ) and in all cases lacked the final oxidation ocess. During operation it could be seen that the blue light iginated in a number of discrete bright points within the ating. The number of these bright points varied considerably om cathode to cathode, some having many hundreds, others erely two or three. The appearance did not remain constant. ividual bright points rapidly appeared and disappeared, and e whole surface often had a 'twinkling' or 'boiling' appearance. he electron current varied with these rapid changes, and a rong hissing sound was heard in the earphones, which was the ason for describing these cathodes as 'noisy'.

In addition to the bright points of light, part or even the whole the oxide coating was covered by a faint general luminescence. his was usually blue, but purple and red luminescences were so observed on some of the cathodes produced by the 'smoke' ethod and processed at temperatures below about 600° C. hese colours tended to disappear and the background lumi-escence to become blue when the cathode was heated or ntinuous current was drawn from the coating.

The ease of starting and the operating stability of the cathodes ere found to depend on the open structure of the positive grid. lany of the emitted electrons are not intercepted directly by this id but pass through to the outer electrodes or the envelope. these are at cathode potential the electrons will be reflected

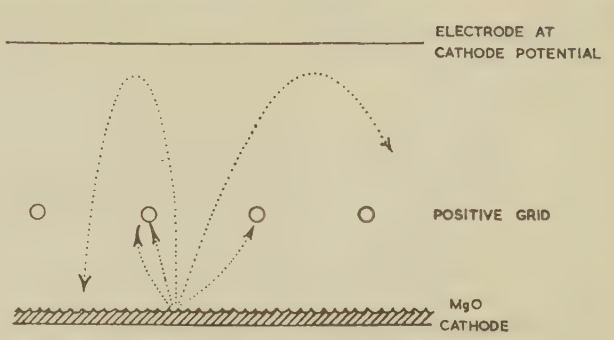


Fig. 4.—Return of electrons to the cathode surface.

back towards the grid, as illustrated in Fig. 4. Again, only a fraction is intercepted. It is conjectured that some electrons have sufficient energy to return to the oxide surface, where they eject several secondary electrons, and so increase the total emission current.

It was found that, if the field beyond the grid does not reflect electrons or if a solid plate anode is used instead of the grid, this return mechanism cannot operate. The returning electrons are responsible for the general blue luminescence which is visible over the whole oxide surface, and for the bright haloes often visible round the bright point emitters. In their absence, neither the background luminescence nor the bright haloes are observed, and the emission is more difficult to start and even less stable during operation.

When the emission was first initiated, an applied potential of between 180 and 300 volts was required to produce a cathode current density of about 1 mA/cm.^{2} The current depended markedly upon the applied potential, but if the mean current density was increased much beyond 2–4 mA/cm.^{2} the emission became unstable. Local particles of the coating became incandescent and emission usually ceased abruptly.

Groups of diodes were placed on life test, with a recording instrument arranged to measure the current through each in turn every two minutes. The magnitudes of the emission currents decreased rapidly from their initial values, and in most cases the current density was only a few microamperes per square centimetre after 100 hours. Parts of two typical traces from the recorder are illustrated in Fig. 5 showing both the decay and the rapid fluctuations of the current.

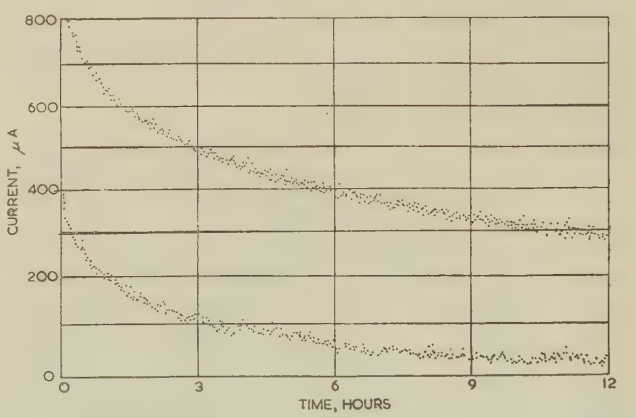


Fig. 5.—Life-test record showing the rapid decay of emission from 'noisy' cathodes.

(3.3) Stable Cathodes

Cathodes processed according to the type of schedule detailed in Section 13.2 behaved in a manner quite different from those

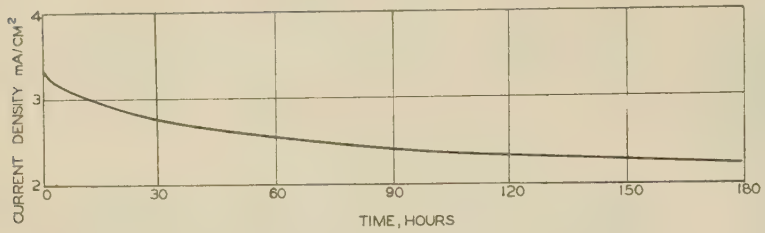


Fig. 6.—The slow decay of emission from a stable cathode.

described above. The emission is no longer restricted to small bright spots but is drawn from the whole surface of the coating. The bright spots, so noticeable on the earlier cathodes, are completely absent. Under a microscope, all the surface particles of the oxide can be seen to be glowing steadily. At low current densities no flickering or rapid fluctuation of the luminescence is observed.

If the cathode coating is too thin, if the surface of the oxide is contaminated, or if the outgassing has been inadequate, the emission may initially be restricted to small areas of the coating. With continued operation such emitting zones spread slowly, and in most cases eventually cover the whole sprayed area.

The life-test records for the fully active cathodes are smooth lines showing a very slow decay. At low current densities (about 0·2 mA/cm^2) this decay is imperceptibly slow and the current record is a straight line. At higher current densities, a slow decay occurs. For example, Fig. 6 shows the change in current for one diode carrying a current density of about 2 mA/cm^2. The record shows very little of the scatter which was a feature of the life-test results for the early type of cathode, and the stability of the current is very much less dependent upon the reflection of electrons back to the oxide surface.

The end of the stable life of the cathode may occur after some thousands of hours, depending upon the current density which has been drawn from the coating. It is usually marked by a growing instability in the life-test record of the current and by the appearance of bright, noisy point emitters on the surface of the oxide.

(4) THE INFLUENCE OF GASEOUS ATMOSPHERE

During the operation of the valve the pressure falls, reaching an equilibrium value of about 10^{-8} torr after some 20–50 hours. This is illustrated in Fig. 7. The pressure was measured by a Bayard-Alpert ionization manometer mounted on the envelope of the diode by a short connecting tube.

To ensure that the observed fall in pressure was not due to the operation of the separate ionization gauge, an experimental magnesium oxide diode was modified to act as an ionization gauge. Two fine probe wires just outside the grid electrode acted as collectors of positive ions. This 'gauge' was first calibrated against the Bayard-Alpert ionization gauge, and the latter was then switched off. The pressure as measured by the ions produced by the magnesium oxide diode fell in the same manner as when measured by the separate ionization gauge. When equilibrium had been reached, interrupting the cold emission caused only a slow and limited increase in pressure. The measurements indicated that with a fully processed cathode the electron emission is accompanied by very little evolution of gas.

The first experiments which showed the effects of injecting more gas into the envelope were carried out with valves containing the early 'noisy' type of cathode. Inside the envelope of one of these valves was mounted a gas generator consisting of

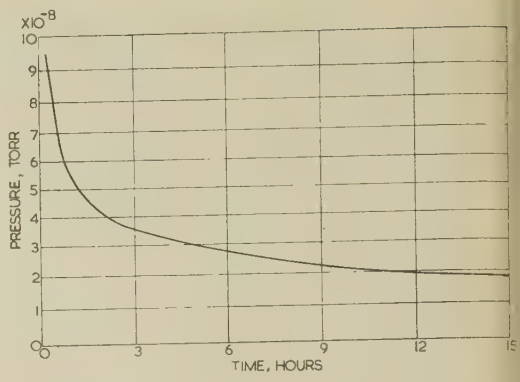


Fig. 7.—Fall in pressure during operation.

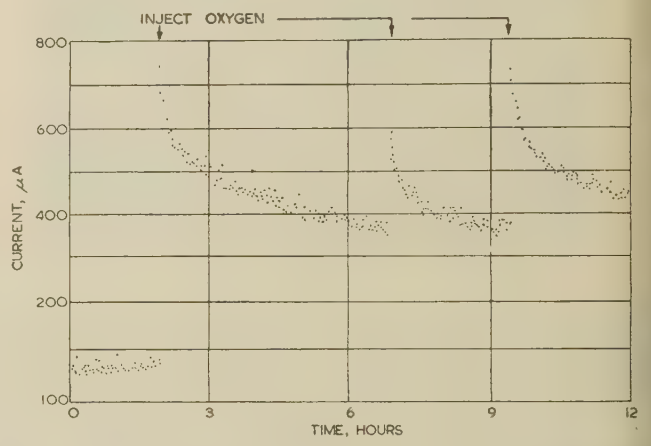


Fig. 8.—Reactivation of a decayed 'noisy' cathode by injecting oxygen.

barium peroxide (in a suitable organic binder) coated on to a tungsten spiral. Such a generator evolves oxygen when heated.*

The life-test record of this valve is illustrated in Fig. 8. This shows that the periodic injection of small amounts of gas, primarily oxygen, reactivated the cathode to its original level. As soon as gas generation ceased, the current decay recommenced. The experiment was repeated with an ionization gauge to monitor the pressure. The increase of current is plotted against the measured pressure in Fig. 9.

The experiment was subsequently repeated with a cathode processed according to the improved schedule, with oxidation at a pressure of 50 torr of air. The valve was run for a sufficient time for the emission to decay from 1·4 to 0·9 mA. The pressure was then increased by heating the barium peroxide. The emission current began to rise at a pressure of about 1×10^{-7} torr. With a recorded pressure of 1×10^{-4} torr the current reached 1·6 mA (higher than its initial value). When injection of gas ceased, the pressure fell rapidly and the slow decay of current recommenced.

Experiments were also carried out with two other types of gas generator, one having a carbonate coating on a tungsten spiral and the other a coating of hydrated alumina. As the tungsten spirals were heated, the pressure rose and some activation of the cathode occurred, but at higher pressures (e.g. 1×10^{-5} torr) the emission steadily deteriorated. In the presence of gas from these two generators the emission became gradually less stable and eventually ceased altogether.

It was also observed that the rate of decay of emission was

* The generator is heated to dull red heat during evacuation of the tube, to eliminate carbonaceous contamination remaining from the binder; otherwise, the oxygen is contaminated with carbon monoxide.

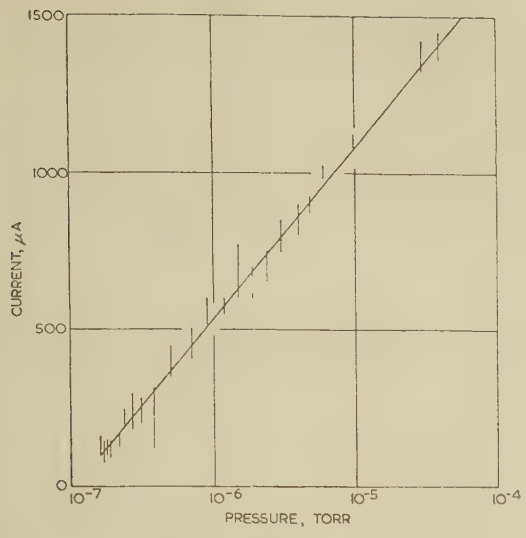


Fig. 9.—Current as a function of pressure during the reactivation of a decayed 'noisy' cathode.

The short vertical lines indicate the extent of current fluctuations.

ower, although the current was less stable, in valves in which ie usual barium getter had been replaced by a relatively inactive lm of evaporated magnesium metal. The results of such simple xperiments as these are not conclusive, although it is clear that he nature of the residual gas atmosphere inside the tube may ave a marked influence on the behaviour of the emitting athode. A much more elaborate vacuum system will be ecessary for a full investigation of the influence of various gases n the emission mechanism. It is nevertheless fairly evident rom the present work that oxygen has an activating effect, and hat water vapour and carbon dioxide (at least at higher ressures) cause a deterioration of the emission mechanism.

(5) VOLTAGE DEPENDANCE OF THE EMISSION

The emission from the magnesium oxide is very voltage ependent. At low current densities the emission followed the aw $J = J_0\varepsilon^{\alpha V_a}$ where V_a is the applied voltage and J_0 is a constant depending upon the condition of the cathode. The easurements were made with a vibrating-capacitor electrometer. As the applied voltage was progressively decreased the current ventually became unstable and the avalanche mechanism failed t an applied potential of about 70 volts. The results of two ypical experiments are shown in Fig. 10. These correspond to alues of J_0 of about $10^{-6}\mu A/cm^2$. The exponential behaviour ontinues up to current densities of about $200\,\mu A/cm^2$. Up to his value the passage of the current through the coating does ot involve the dissipation of much power, and there is little ncrease in the temperature of the cathode. If the current is ncreased further, resistive heating raises the temperature of he coating—whose resistivity consequently decreases. The oltage/current relationship is thereafter controlled by the emperature of the cathode.

(6) CATHODE TEMPERATURE

(6.1) Temperature as a Function of Current

As the electrons pass through the oxide coating they lose nergy to the particles of oxide and the temperature of the whole athode rises. This temperature was measured with the aid of platinum/13% rhodium-platinum thermocouple. Two very ine thermocouple tapes were welded to the nickel core, just at he end of the coated area. These tapes were brought out

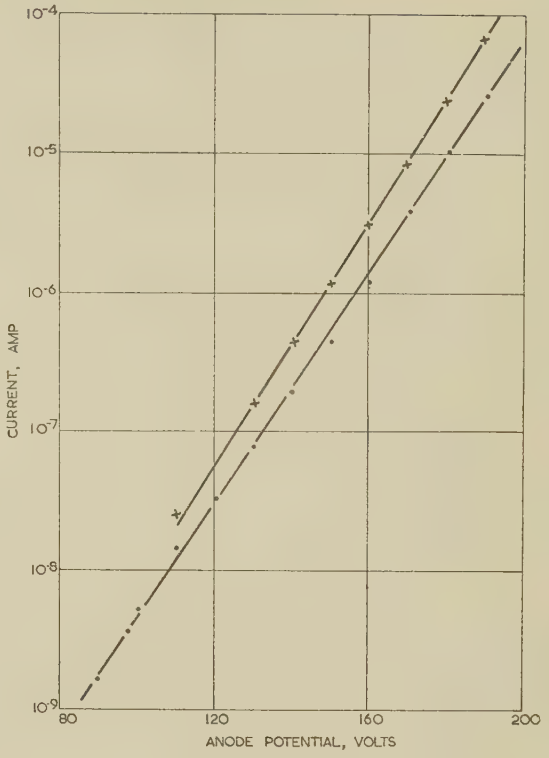


Fig. 10.—Voltage dependence of the emission.

through holes in the mica support to be joined to heavier platinum and rhodium-platinum wires sealed directly through the glass envelope. The thermo-potential was measured with a moving-coil instrument calibrated directly in temperature. The measurements are illustrated in Fig. 11. The lowest line is

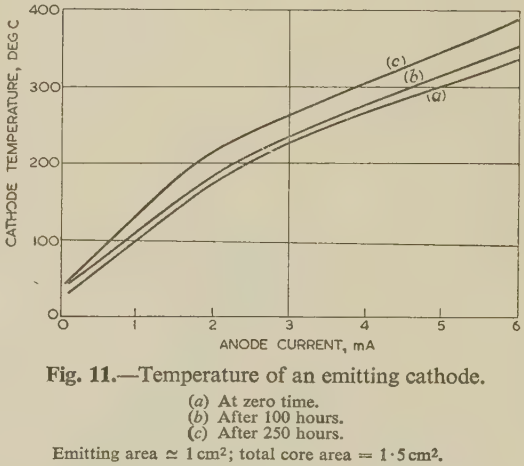


Fig. 11.—Temperature of an emitting cathode.

(*a*) At zero time.
(*b*) After 100 hours.
(*c*) After 250 hours.

Emitting area ≃ 1 cm²; total core area = 1·5 cm².

for a newly processed cathode and the upper lines are for the same cathode after some hours of operation at approximately $1{\cdot}5\,mA/cm^2$.

Interesting information can be derived from these data if it is assumed that the cathode temperature for a given power dissipation is the same whether the heat is generated in the coating or by the internal heater. (Provided that the temperature is high enough for conduction losses to be small compared with the radiation loss, this is a reasonable approximation.) Thus, temperatures quoted in Fig. 11 can be converted into values of power dissipated. If it is now assumed that the cathode coating

is equivalent to a resistance R_x carrying the anode current i_a with potential drop of $V_x = i_a R_x$ across the resistance, it is possible to calculate values for R_x and V_x.

On this basis, the calculated equivalent potential of the oxide surface, V_x, is found to vary very little for current densities between 0·5 and 8 mA/cm^2, being comparable with the anode potential, V_a. Some experimental values are given in Table 1.

Table 1

CALCULATED VALUES OF CATHODE EQUIVALENT SURFACE POTENTIAL AND EQUIVALENT RESISTANCE

Current i_a	Temperature	Calculated V_x	Calculated R_x
μA	deg. C	volts	Kilohms
500	71	200	404
1 130	141	186	165
1 930	213	200	104
2 700	259	209	77
3 700	292	189	51
4 850	348	212	44
5 750	381	223	39

The emitting area of the cathode in this experiment was about 1 cm^2, but the temperature of the cathode depends upon radiation from the whole area of the core, coated and uncoated (1·5 cm^2). As the cathode ages the value of V_x increases, as shown in Fig.12.

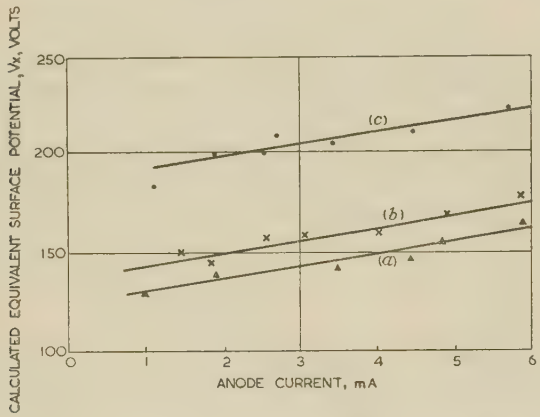


Fig. 12.—Calculated equivalent potential, V_x, of the oxide surface.
(*a*) At zero time.
(*b*) After 100 hours.
(*c*) After 250 hours.

Steady currents greater than those shown in Table 1 cause higher cathode temperatures, but the emission tends to become unstable when the temperature approaches 400° C. It is, however, easy to draw much larger currents for brief periods without overheating the cathode. The maximum current density so far observed (from a coating of restricted area) is 300 mA/cm^2, but this could not be maintained for more than a few seconds because of the rapid rise in temperature.

(6.2) Current as a Function of Temperature

It has been suggested above that the magnesium-oxide cathode can produce high levels of emission only for brief periods, because the emission mechanism fails as the temperature of the coating is increased. This failure of the emission mechanism has also been demonstrated in life-test experiments, the results of one experiment being illustrated in Fig. 13. With a constant anode potential (290 volts) the power supplied to the heater inside the cathode is progressively increased. As the temperature of the cathode rises, the emission first increases slightly and then begins to decay. The blue luminescence on the coating becomes progressively less stable and begins to flicker. The emission current becomes very erratic and eventually ceases altogether. However, recovery is very rapid if the cathode is allowed to cool before emission has ceased completely.

If the heating of the cathode is continued to temperatures higher than about 700° C, the stable emission mechanism may

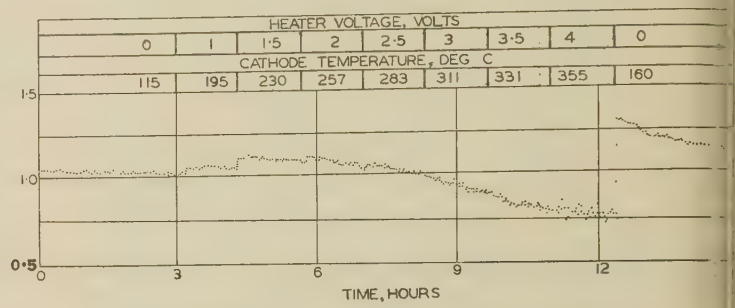


Fig. 13.—Effect of heating the cathode during life-test.

be permanently damaged. Subsequently, the cathode may either fail to produce any self-sustained emission, or may operate as a 'noisy' cathode with a limited number of bright emitting points.

(7) ENERGY DISTRIBUTION OF THE ELECTRONS

One method of measuring the energy of electrons is by determining the height of an electrostatic barrier they are just able to surmount. A simple measurement of this kind is possible with the triode assemblies already described. In the normal life-tests the outer grid was operated at cathode potential. If this grid is now made progressively more positive, the height of the potential barrier which faces electrons leaving the inner grid is progressively reduced.

When the barrier becomes sufficiently low, some electrons reach the second grid, and as the barrier is reduced further, the current flow to the outer grid increases. Fig. 14 shows the ratio

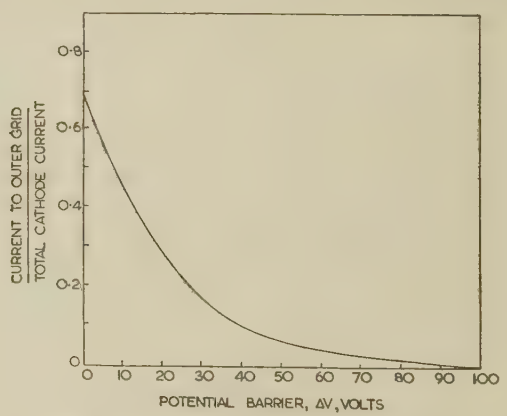


Fig. 14.—A retarding field characteristic.

between the current flowing to the outer grid and the total cathode current plotted against the height of the potential barrier ΔV. (ΔV represents the difference in potential between the inner and outer grids of the triode structure.) Similar curves were obtained for a number of valves under various

nditions. It is interesting to note that a few electrons have an nergy exceeding 100 volts, whereas about 10% have a kinetic nergy greater than 40 volts when leaving the first grid.

This wide range of energy is partly due to scattering of the ectrons as they pass through the first grid. However, the otential of the second grid directly affects the emission, both by anging the potential gradient above the cathode surface and preventing electrons from returning to the oxide surface. he curve in Fig. 14 and the histogram (Fig. 15) derived from it

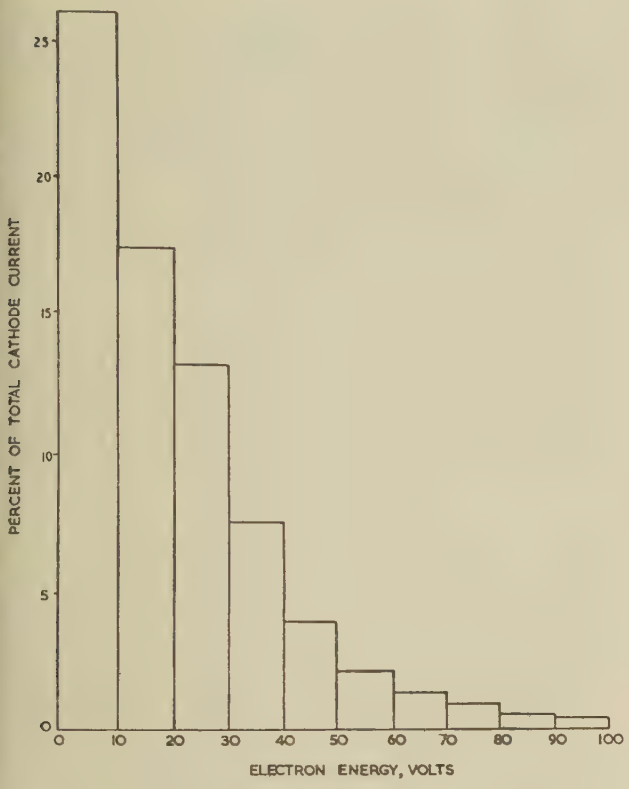


Fig. 15.—Approximate distribution of electron energies.

herefore represent only approximately the energy distribution of the electrons actually emitted by the magnesium oxide. These Figures do, however, indicate the wide spread of energies which would be available if the simple diode were used as a ource of electrons.

An energy distribution such as that in Fig. 15 corresponds pproximately to the Maxwellian distribution for electrons mitted thermionically at about 200 000° K.

(8) EXPERIMENTS ON REPROCESSING

At high current densities the emission from magnesium oxide is impaired by the increased temperature; some form of cooling is therefore desirable. A possible solution to the problem is to process the valve in the normal way, and then break open the envelope and remount the electrode structure. The heater is replaced by a degassed copper mandrel, to remove heat from the cathode. The new valve is then evacuated. This type of experiment has been only partially successful. Some contamination of the oxide occurs during the remounting and reprocessing, and stable emission is not usually possible without further outgassing.

A partial solution which has been more successful is to increase the radiation and conduction cooling of the cathode while leaving a heater inside the core. In practice, the amount of cooling achieved is limited by the need to avoid overloading this tungsten heater during processing.

(9) DISCUSSION

The emission of large numbers of electrons from a magnesium-oxide cathode depends upon the avalanche multiplication of a very small primary current. Jacobs[6] has shown that it is unlikely that this primary current is produced by a simple Fowler-Nordheim cold-emission mechanism, but that it may be photo-electric in origin. The avalanche process is pictured in the following simple description. The current multiplication may take place both inside and outside the particles of magnesium oxide. As the primary electrons pass through the oxide under the influence of the steep field gradient they will acquire sufficient kinetic energy to produce pairs of electrons and holes. These in turn will accelerate and eventually produce further pairs.

If the electron affinity (i.e. the external work-function) at the surface of the oxide is small compared with the energy necessary to create an electron-hole pair, some of the energetic electrons will be able to escape from the surface of the oxide into the open pores. (This process is analogous to that by which energetic electrons are emitted from semiconductor junctions.[15])

Once the electrons escape from the surface of the oxide they are accelerated through the porous matrix, but whenever they strike the surface of the oxide they will cause the release of further electrons by secondary emission. The condition for a high yield of secondary electrons is also that the electron affinity should be relatively small compared with the width of the energy gap in the electronic band structure of the oxide.

The band structure of magnesium oxide is very complex, with several donor and acceptor levels determined by the presence of excess magnesium or oxygen in the crystal lattice. It is conjectured that it is the energy levels produced by excess oxygen in the lattice which are essential for stable emission. This would explain the necessity for oxidation during processing and the recovery of emission which occurs when oxygen is injected.

During operation, the surfaces of the oxide particles are continuously bombarded by electrons as they pass outwards through the pores of the matrix. It is suggested that this bombardment causes some dissociation of the oxide with the partial elimination of the excess oxygen levels and consequent reduction in the efficiency of the avalanche mechanisms.

The experimental evidence strongly suggests that the rate of decay depends upon the local density of the electron current and upon the steepness of the potential gradient through the coating. The decay of emission is therefore minimized by a cathode matrix structure such that the total current is evenly distributed and the mean free path of the electrons in the pores is kept short. This will reduce the energy of the electrons which bombard the oxide surface, and so reduce the rate of dissociation of the oxide.

(10) SUMMARY AND CONCLUSIONS

In this study of the magnesium-oxide cathode the chemical nature of the (fairly pure) starting compounds has been found to be of little importance compared with the structure and state of oxidation of the finished coating of magnesium oxide. In the present stage of development the processing of the cathode is relatively slow, involving the injection of air into the vacuum system. It has not been found possible to carry out this oxidation as part of the component pre-processing before assembly, because the oxide is extremely susceptible to subsequent surface contamination, which can be removed only by high-temperature decomposition and oxidation.

An open-structured coating with a thickness of 60–80 μ has

been found to give the best results; this appears to conflict with findings of the American workers, who have preferred 30–40 μ. However, it seems probable that the optimum thickness will be influenced by particle size and pore dimensions, and these aspects have not been quantitatively studied in the present work.

Given a coating in which the important properties of solid resistivity of the magnesium oxide and the average path length for free flight of electrons are more or less uniform, the magnesium-oxide cathode is capable of remarkable stability under conditions of low current density.

Although it has been referred to as a 'cold' cathode, in fact there is a distinct dissipation of power as heat in the coating, and special cooling is necessary to permit maintenance of steady emission greater than 0·5–1·0 mA/cm^2.

The emission current is not fully space-charge smoothed and the available electrons have been shown to possess a wide range of energies. One result of this is that the current is noisier and more difficult to control than a similar current from a conventional cathode.

Considering the usefulness of this type of cathode in electron devices, the absence of heater current supplies and the lower operating temperature are of major importance. The fact that the cathode operates satisfactorily in the presence of relatively higher pressures of oxidizing gases than obtain in the normal vacuum valve might also be of value in special applications. The property of passing a very small stable current with the dissipation of very little power combined with the ability to pass short bursts of high current density suggests possible applications where a vacuum device with magnesium-oxide cathode may be more suitable than either a thermionic or solid-state device.

The need to have a positive field gradient at the cathode surface, the need for separate current-initiation and the wide energy spread of the emitted electrons constitute special features which place limitations on the design and employment of devices incorporating this type of cathode.

(11) ACKNOWLEDGMENTS

Acknowledgment is made to the Engineer-in-Chief of the Post Office for permission to make use of the information contained in the paper. The author also wishes to acknowledge experimental contributions by Mr. R. E. Hines and helpful discussions with Mr. E. F. Rickard during the course of the work.

(12) REFERENCES

(1) Malter, L.: 'Thin Film Field Emission', *Physical Review*, 1936, **50**, p. 48.

(2) Koller, L. R., and Johnson, R. P.: 'Visual Observations on the Malter Effect', *ibid.*, 1937, **52**, p. 519.

(3) Nelson, H.: 'Phenomenon of Secondary Emission', *ibid.*, 1939, **55**, p. 985.

(4) Nelson, H.: 'Field-Enhanced Secondary Electron Emission', *ibid.*, 1940, **57**, p. 560.

(5) Dobischek, D., Freely, J. B., and Jacobs, H.: 'Dynode Coating', United States Patent No. 2802127.

(6) Dobischek, D., Jacobs, H., and Freely, J.: 'The Mechanism of Self-Sustained Electron Emission from Magnesium Oxide', *Physical Review*, 1955, **91**, p. 804.

(7) Dobischek, D., and Jacobs, H.: 'Cold-Cathode Vacuum Tube', United States Patent No. 2842706.

(8) Dobischek, D., Schweitzer, J. A., and Ward, P. T.: 'Method of Making an Electron Emitter', United States Patent No. 2873218.

(9) Jacobs, H.: 'Field Dependent Secondary Emission', *Physical Review*, 1952, **84**, p. 877.

(10) Jacobs, H., Freely, J., and Brand, F. A.: 'The Mechanism of Field-Dependent Secondary Emission', *ibid.*, 1952, **88**, p. 492.

(11) Veenemans, C. F., and van Zanten, P. G.: 'Cold Cathode Discharge Tube', United States Patent No. 2491234. (See also French Patent No. 920273 and Dutch Patent No. 65336.)

(12) Skellett, A. M., Firth, B. C., and Mayer, D. W.: 'The Magnesium Oxide Cold Cathode and its Application in Vacuum Tubes', *Proceedings of the Institute of Radio Engineers*, 1959, **47**, p. 1704.

(13) Holmes, M. F., and Reynolds, F. H.: 'Post Office Valves for Deep-Water Submarine Telephone Repeaters', *Proceedings I.E.E.*, Paper No. 3163 E, March, 1960 (**107 B** p. 165).

(14) Woods, J., and Wright, D. A.: 'Field-Enhanced Cathodoluminescence', *Proceedings of the Physical Society, B* 1955, **68**, p. 566.

(15) Burton, J. A.: 'Electron Emission from Avalanche Breakdown in Silicon', *Physical Review*, 1957, **108**, p. 1342.

(13) APPENDICES

(13.1) Spray Formulations

Reagent grade materials were used throughout.

(*a*) Magnesium oxide (heavy) . . . . . 16 g
Magnesium carbonate (heavy) . . . . . 12 g
Magnesium oxalate . . . . . 12 g
Roll in 200 cm^3 of a 9 : 1 mixture of *n*-propanol and amyl acetate, dilute with the same mixture to 600 cm^3 and add a small quantity of poly-*n*-butyl methacrylate as binder.

(*b*) Magnesium peroxide (containing 25% peroxide) . 22 g
Magnesium carbonate (heavy) . . . . . 20 g
Roll in 300 cm^3 of amyl acetate and dilute with 300 cm^3 of *n*-propanol; no binder is needed.

(13.2) Processing Schedule

Evacuate to a pressure of 10^{-5}–10^{-6} torr.
Bake for 1 hour at 350° C.
Cool the diffusion pumps, allowing the pressure to rise to about 40 millitorr.
Decompose the cathode coating by heating for 2 min at 850° C.
Isolate the manifold from the pumping system and inject air to a pressure of 50–60 torr.
Oxidize the cathode by heating for 5 min at 700° C.
Re-evacuate to 10^{-6} torr.
Rebake for 20 min at 350° C.
Outgas the cathode by heating for 5 min at 620° C.
Flash the getters and seal.

621.317.794.029.6: 621.375.9: 539.19

The Institution of Electrical Engineers
Paper No. 3331 E
Jan. 1961

MICROWAVE SPECTROSCOPY

By D. J. MILLEN, B.Sc., Ph.D.

(*The paper was first received 14th April, and in revised form 5th July,* 1960.)

SUMMARY

The scope of microwave spectroscopy and its relationship to other branches of spectroscopy are briefly indicated, attention being especially directed to those aspects which are important in the design of spectrometers. The factors determining spectrometer sensitivity are discussed. The basis of molecular modulation is presented with special reference to Stark modulation. A survey is made of the various types of spectrometer and their relative merits are discussed. The application of the method to chemical analysis is considered and indications are given of the directions in which further experimental work is required.

LIST OF SYMBOLS

A = Einstein's coefficient of spontaneous emission.
α = Absorption coefficient.
B = Einstein's coefficients of absorption and induced emission.
f = Frequency.
h = Planck's constant.
I = Moment of inertia.
J = Quantum number giving total angular momentum in units of $h/2\pi$.
K = Quanta of angular momenta along a molecular axis (units of $h/2\pi$).
k = Boltzmann's constant.
L = Conversion loss.
M_J = Quanta of angular momenta along a space-fixed axis (units of $h/2\pi$).
N = Noise figure.
P = Power.
ρ = Energy density of radiation.
T = Temperature.
W = Energy.
ψ = Wave-functions.
V = Voltage.
v = Vibrational quantum number.
Z = Impedance.

(1) MOLECULAR SPECTRA

The development of microwave techniques has made possible the opening up of a new branch of experimental spectroscopy, which has yielded new and more precise information about molecules and to some extent atoms. The origin of molecular spectra may be understood in terms of the general pattern of molecular energy levels shown in Fig. 1. The quantized energy levels W_1 and W_2 corresponding to electronic motion are widely spaced, while the molecular vibrational energy levels W_v are much more closely spaced. The levels W_J associated with molecular rotation are even more closely spaced. Transitions between a pair of quantum levels with an energy separation of ΔW may under certain conditions be brought about by the absorption of radiation of frequency f according to the Bohr frequency condition $\Delta W = hf$, where h is Planck's constant.

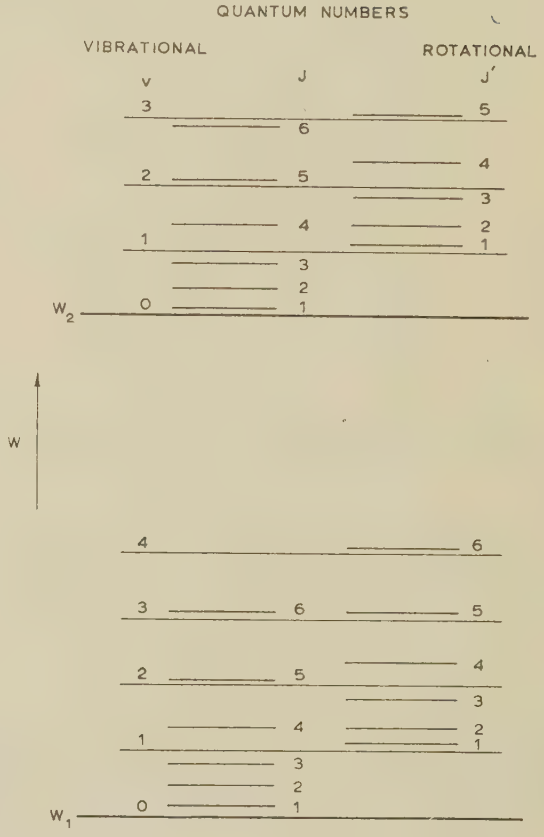


Fig. 1.—The general pattern of molecular energy levels.

Transitions between the rotational levels, W_J, lead to absorption in the microwave region.
Vibrational transitions occur in the infra-red, and electronic spectra in the visible or ultra-violet regions.

The wide spacing of the electronic levels leads to absorption in the ultra-violet region of the spectrum. Vibrational spectra occur at lower frequencies in the infra-red region, and rotational changes involving still smaller energies lead to absorption for almost all molecules in the microwave region. The development of microwave spectroscopy has made possible the direct observation of rotational spectra. Hitherto almost all information about molecular rotation had to be derived from the study of fine structure in other types of spectra. The direct study has numerous advantages, and has in turn yielded an interesting hyperfine structure arising through the coupling of nuclear spin and molecular rotational momenta.

(1.1) Rotational Spectra

The rotational energy levels of a linear molecule, regarded as a rigid rotor, are given in terms of a quantum number $J = 0, 1, 2, \ldots$ by

$$W_J = \frac{h^2}{8\pi^2 I} J(J+1) \quad . \quad . \quad . \quad . \quad . \quad (1)$$

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
This is an 'integrating' paper. Members are invited to submit papers in this category, giving the full perspective of the developments leading to the present practice in a particular part of one of the branches of electrical science.
Dr. Millen is Reader in Chemistry, Sir William Ramsay and Ralph Forster Laboratories, University College, London.

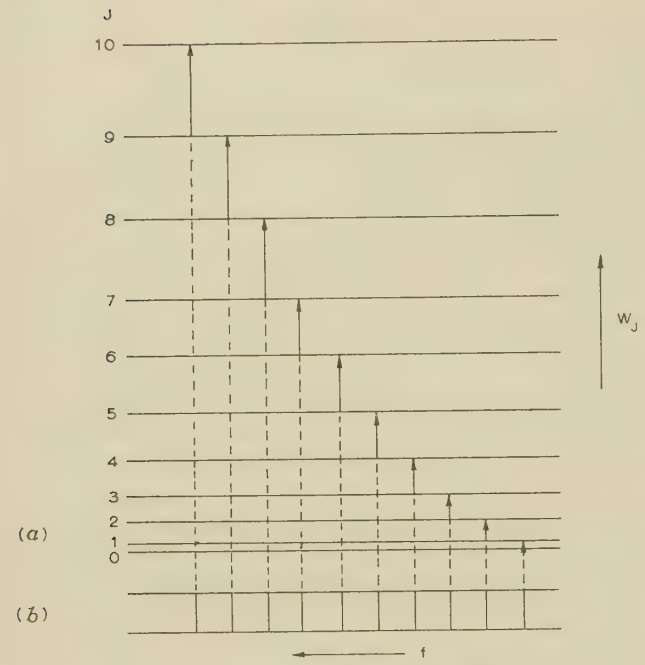


Fig. 2.—Quantum levels for a linear rigid rotor.

The possible transitions as a result of absorption of dipole radiation are connected by tie-lines to frequencies in the spectrum.

where I is the molecular moment of inertia. The transitions which may be brought about by absorption of radiation and the resulting spectra are indicated in Fig. 2. Observed spectra agree closely with this pattern, although for precise agreement it is necessary to take into account centrifugal distortion.

Non-linear molecules fall into two classes. For symmetric rotors, e.g. CH_3F, with two equal principal moments of inertia, the energy levels are rather more complicated, but the spectra have the same simplicity as for linear molecules. For asymmetric rotors, e.g. H_2O, in which all three principal moments differ, the energy levels and the spectra are quite complicated, and it is, in general, a major problem to interpret the spectrum. Fig. 3 illustrates the difference between the two cases, showing

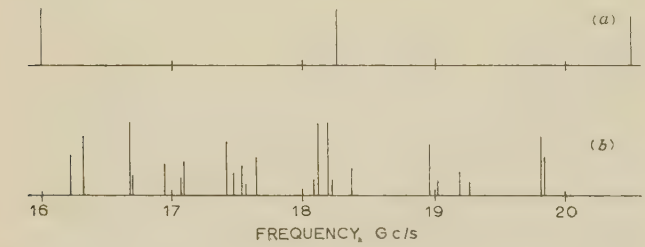


Fig. 3.—Comparison of absorption spectra in the range 16–20 Gc/s.

(*a*) a symmetric rotor, CH_3HgBr.
(*b*) an asymmetric rotor, HNO_3.

The spectrum in each case is for a single isotopic species and effects due to centrifugal distortion and hyperfine structure are not included.

the spectra observed[1,2] for CH_3HgBr and HNO_3 over a range of 4 Gc/s.

In either case, analysis of the spectrum yields the molecular moments of inertia, and from these, inter-atomic distances and bond angles within the molecule may be obtained. Where there are several structural parameters to be determined, the additional information may be obtained from the spectra of isotopically substituted molecules, e.g. DNO_3, $H^{15}NO_3$, etc., for which bond distances remain unchanged but which have different moments of inertia.

Because of the high accuracy of the frequency measurements, molecular structures may be obtained with great precision. For diatomic molecules the accuracy of the bond distance is set by the present accuracy of Planck's constant; for polyatomic molecules the accuracy is often limited by the difficulty of taking into account the zero-point internal motion of atoms within molecules. Nevertheless, bond distances can often be specified to within a few thousandths of an Ängstrom unit.

(1.2) The Stark Effect

The Stark effect has proved to be of great importance in both the theoretical interpretation of microwave spectra and the design of spectrometers. The effect to be discussed is that of an applied external electrical field on the rotational energy levels and the microwave spectrum of a molecule. The effect takes its name from Stark, who discovered it in atomic spectroscopy. Its origin may be understood by considering a symmetric rotor molecule such as CH_3F in a uniform electric field. For a molecule possessing J quanta of angular momenta, space quantization of the angular momentum along the field direction leads to a splitting of the level into $2J + 1$ states. Each of these corresponds to a particular quantum value, M_J, of the component of angular momentum along the field direction, and clearly M_J may have any one of the $2J + 1$ quantum values $J, J - 1, \ldots 0 - (J - 1), -J$. The splitting and the spectroscopic consequences are illustrated in Fig. 4. A state characterized by a

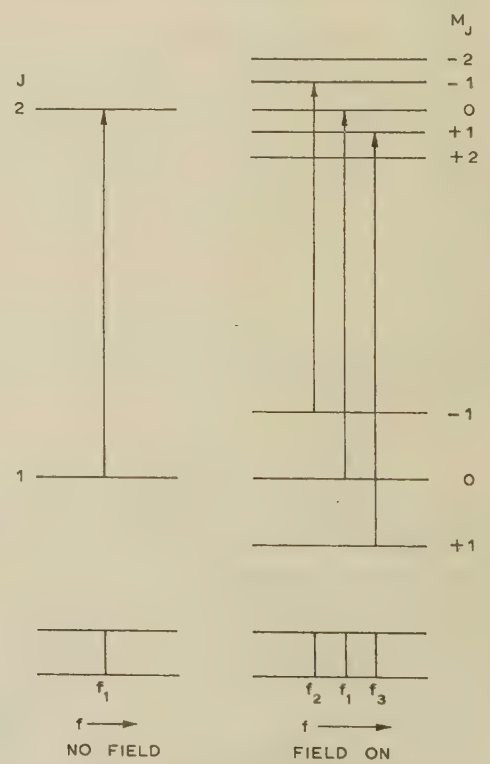


Fig. 4.—The Stark effect.

The splitting of the $J = 1$ and $J = 2$ quantum levels of a symmetric rotor and its effect on the spectrum are shown under the usual conditions for microwave propagation, for which dipole transitions must satisfy the condition $\Delta M_J = 0$.

particular component of momentum M_J along the field direction also has a particular value for the component in the field direction of the molecular electric dipole moment. Thus there is a distinct

teraction energy for each of the $2J + 1$ states; this is given in general by

$$W_{JKM_J} = -\frac{\mu E K M_J}{J(J+1)} \quad . \quad . \quad . \quad . \quad (2)$$

where E is the external electric field. Two valuable consequences can be seen immediately: first, eqn. (2) allows the determination of the molecular dipole moment from the Stark splittings; secondly, the number of components into which a given line is split is seen to depend on the J-values of the levels involved. This effect is a valuable aid in the interpretation of the spectra of symmetric rotors where J Stark components for a line indicates that at one of the levels involved in the transition has quantum number J. The way in which the effect is used[2] is shown in Fig. 5. The application of the effect in spectrometer design is discussed in Section 3.1.

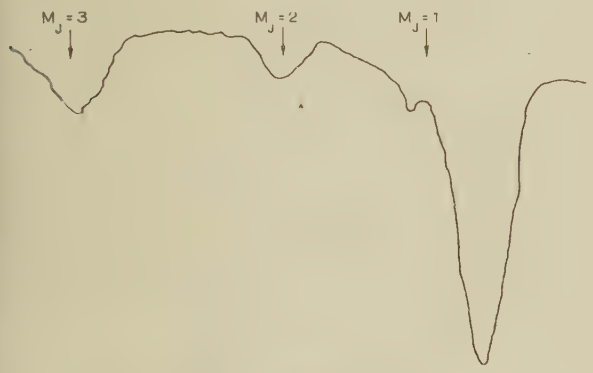


Fig. 5.—The Stark effect.

On the right is the field-free line; the three Stark components, $M_J = 1, 2, 3$, indicate that for one level involved in the transition $J = 3$, and the relative intensities show whether that the total angular momentum is unchanged for the transition. The line illustrated here occurs at 22 161·00 Mc/s in the spectrum of HNO_3.

(1.3) Inversion Spectra

In addition to overall rotation of molecules, both internal rotation within molecules and molecular inversion have consequences which are frequently detectable in the microwave region. The inversion of ammonia will be discussed briefly, since it has been exploited in the development of gaseous masers discussed in Section 4.3. The method has yielded exceptionally high spectroscopic resolution.

Two equivalent configurations may be proposed for a pyramidal NH_3 molecule, as shown by the broken lines in Fig. 6. The relation between these is such that, if the three hydrogen atoms were replaced by distinguishable groups, the two configurations would represent mirror-image optical isomers. The upper curve shows the potential energy of the system as a function of the pyramidal height d, the pyramidal configuration being more stable than the planar configuration. If the barrier between the two minima approached infinity, as shown in the lower curves, two sets of energy levels would result, giving a pair of levels for each allowed energy value with wave-functions ψ_0^a, ψ_0^b; ψ_1^a, ψ_1^b, etc. A consequence of a finite barrier is a splitting of these energy states into two closely lying states, as shown in Fig. 6(*a*), the wave-function being represented as linear combinations

$$\psi_0 = \psi_0^a + \psi_0^b$$
$$\psi_0' = \psi_0^a - \psi_0^b$$

and similarly for levels 1, 2, 3, 4, ... The extent of the splitting is a sensitive function of the height and width of the potential barrier. Application of the Bohr frequency condition $\Delta W = hf$ shows that the transition between the pair of levels, as indicated in Fig. 6, may be brought about by the absorption of radiation

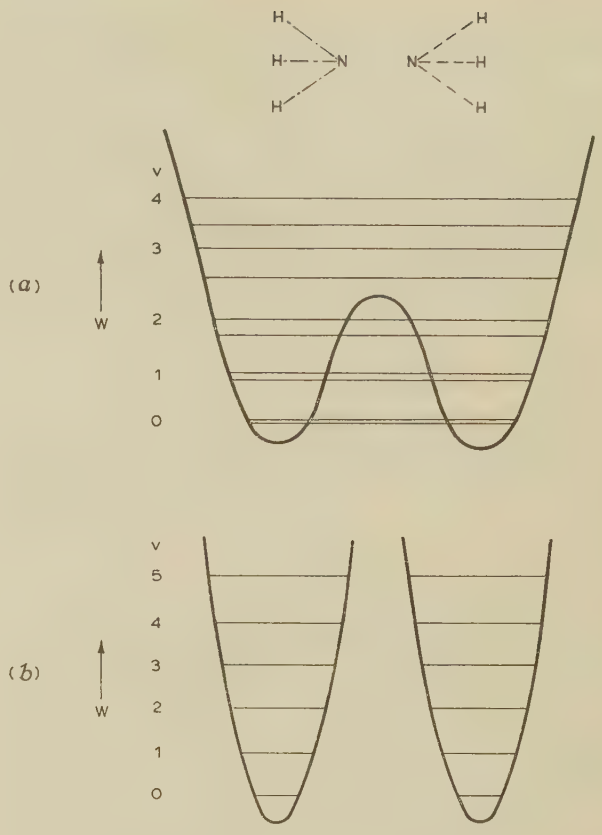


Fig. 6.—The inversion of ammonia.

Transitions between the components of the $V = 0$ level in the upper potential curve give rise to absorption in K-band.

(*a*) Low-barrier case.
(*b*) Infinite-barrier case.

of a frequency which falls in the K-band. Actually there is found a transition of this kind for each rotational state of the molecule. This may be understood by considering the rotation to produce a centrifugal distortion of the distance d, and so the effective barrier and consequently the splitting is dependent on the rotational state.

The absorption of radiation is not to be regarded as bringing about inversion of configuration. It leads to a transition between two energy states, for either of which examination of ψ^2 and ψ'^2 shows that there is an equal probability of finding the NH_3 molecule in either configuration.

Inversion has an important consequence for the Stark effect of ammonia. The first-order interaction μE averaged over an inversion-rotation state is zero. To obtain the interaction energy it is necessary to consider second-order perturbation;[3] the dipole-field interaction perturbs the motion from its field-free form in such a way that there now results an interaction energy dependent on $\mu^2 E^2$. As usual in such perturbations, the perturbed energy levels move away from each other. The result is illustrated in Fig. 7 for the rotational state $J = 3$, $K = 3$ in each inversion state. The effect, which contrasts with the normal effect for a symmetric rotor (Fig. 4), is exploited in the gaseous maser.

(1.4) Intensities of Absorption Lines

The factors which determine the intensities of absorption lines are important in considerations of both spectrometer and maser design. Consider an assembly of molecules in the presence of monochromatic radiation with energy density ρ and frequency f

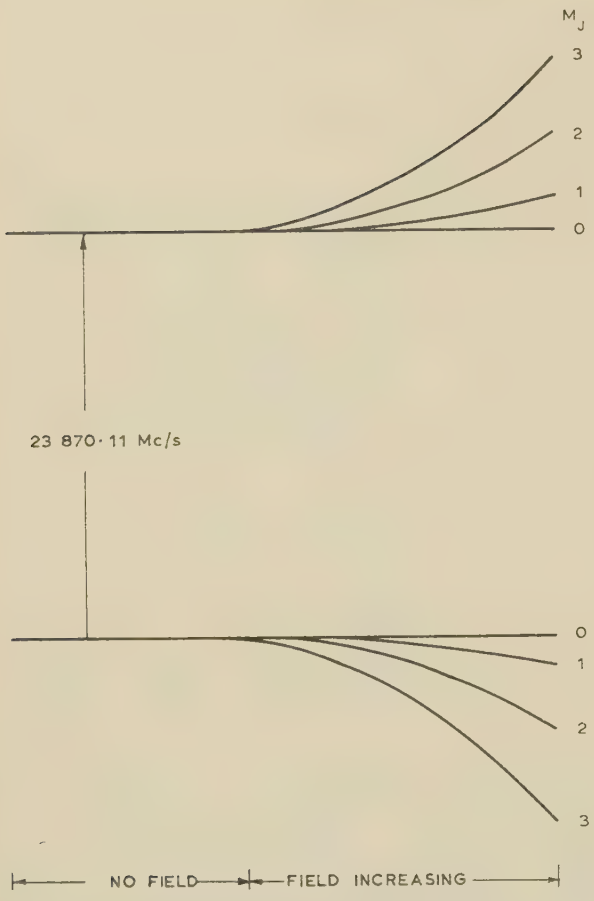


Fig. 7.—Effect of electric field on the rotational level $J = 3$, $K = 3$ for the upper and lower inversion states of ammonia.

such that $f = (W_m - W_n)/h$ where W_m and W_n are two energy levels between which a transition is allowed, as represented for various pairs of levels in Fig. 2. There are three processes to be considered: absorption, $W_n \rightarrow W_m$; induced emission, $W_m \rightarrow W_n$; and spontaneous emission, $W_m \rightarrow W_n$. The probability of the first process per molecule, in time dt, is $B_{nm}\rho dt$, and that of the second is $B_{mn}\rho dt$; the probability, $A_{mn}dt$, of the last process is independent of the density of the radiation introduced. The coefficients B_{nm}, B_{mn} and A_{mn} are known as Einstein's coefficients of absorption, induced emission and spontaneous emission, respectively. By thermodynamic arguments and the use of the black-body radiation law, Einstein showed that

$$\left.\begin{aligned} B_{mn} &= B_{nm} \\ \text{and} \quad A_{mn} &= \frac{8\pi hf^3}{c^3} B_{mn} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (3)$$

For microwave frequencies $A_{mn} \ll B_{mn}$ and for many purposes it may be neglected.

The observed result of introducing microwave radiation into the assembly of molecules is thus the net effect of absorption and induced emission. If the molecules are in thermal equilibrium at temperature T, the number of molecules, N_n and N_m, in the non-degenerate levels W_n and W_m are related by Boltzmann's law, i.e.

$$\frac{N_m}{N_n} = \exp - \frac{W_m - W_n}{kT} \quad . \quad . \quad . \quad . \quad (4)$$

The net number of molecules undergoing absorption in time dt is $(N_n - N_m)B_{nm}\rho dt$, and the net decrease in energy density is

$$d\rho = hf(N_n - N_m)B_{nm}\rho dt \quad . \quad . \quad . \quad . \quad (5)$$

This quantity is closely related to the absorption coefficient which is defined as

$$-\frac{1}{P}\frac{dP}{dx}$$

where P is the power being propagated through the gas in the x-direction. Since $P \propto \rho$ and $dx = cdt$, it follows that the absorption coefficient is given by

$$\alpha = \frac{hf}{c}(N_n - N_m)B_{nm} \quad . \quad . \quad . \quad . \quad . \quad (6)$$

A number of consequences emerge. First, the effect of induced emission on the intensities of microwave absorption lines is serious, for taking $f = 30\,\text{Gc/s}$ and $T = 300°\,\text{K}$ gives $N_m/N_n = 0{\cdot}995$. The result is that, even for strong absorption lines, α does not exceed about $10^{-3}\,\text{cm}^{-1}$ under optimum pressure and for many substances the values are several orders of magnitude smaller. High sensitivity is an important consideration in spectrometer design. Secondly, the absorption of microwave power will increase the population N_m of the upper state at the expense of the lower state, and if thermal relaxation is sufficiently slow, the intensity of the peak of the absorption line may fall with increasing power. This saturation effect frequently limits the amount of microwave power which may be profitably used. Finally, where a method can be devised to increase N_m such that $N_m > N_n$, the introduction of radiation into the system will lead to net emission and there arises the possibility of amplification at the frequency f.

The discussion above neglects the finite width of the absorption line arising from a number of causes discussed in Section 4.2. These may be taken into account by introducing a line shape factor $S(f)$ into eqn. (6). The conclusions remain as above.

(2) EXPERIMENTAL CONSIDERATIONS

The attenuation of microwave power by gases at atmospheric pressure varies only slowly with frequency and sharp spectral absorption lines are not observed. This arises because under these conditions molecules are subjected to frequent intermolecular collisions and the lifetime of a rotational state is short. Sharp absorption lines are obtained only by working at low pressures (generally 10^{-1} to $10^{-3}\,\text{mm Hg}$). At the lower end of this range the intensity of a line peak increases with pressure; towards the upper end of the range a pressure increase produces no further increase in peak intensity, but leads to increased total absorption of radiation by increase of the line width. Under optimum pressure the attenuation produced by the strongest lines is about $10^{-3}\,\text{dB/cm}$, but many substances produce spectral lines having much smaller attenuations and a general-purpose spectrometer normally needs to detect lines with a peak absorption coefficient, α, down to about $10^{-8}\,\text{cm}^{-1}$.

(2.1) Spectrometer Sensitivity

Considerations of the sensitivity[4–6] of a spectrometer may be compared with those made in the treatment of the sensitivity of a radar receiver. In order to do this it must be remembered that there is one very important difference: in a radar receiver the power to be detected consists of a weak signal, δP, frequently in the form of a pulse; in microwave spectroscopy the effective signal is a small decrement, ΔP, in a much larger amount of microwave power, P. This decrease, ΔP, is equivalent, not to a

lse of power ΔP, but, as shown below, to a much smaller antity $(\Delta P)^2/4P$.

Let Z be the impedance of a waveguide through which power lows, where

$$V^2 = ZP$$

A decrement in power ΔP gives rise to a voltage decrement ΔV_s, available as a signal, which is given by

$$\Delta V_s = \frac{Z\Delta P}{2V} = \frac{\Delta P}{2\sqrt{(ZP)}} \quad . \quad . \quad . \quad . \quad (7)$$

ΔP is sufficiently small for the terms in $(\Delta V_s)^2$ to be neglected. In the case of linear detection, which is appropriate to superheterodyne and various modulation spectrometers to be described later, the output voltage of an amplifier of gain G will be

$$V_0 = G\Delta V_s \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

The thermal noise to be considered arising in a source of impedance Z is

$$P_n = kT\Delta f \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

where Δf is the bandwidth of the amplifier. The noise output voltage, V_n, of the amplifier is

$$V_n = G(ZNkT\Delta f)^{1/2} \quad . \quad . \quad . \quad . \quad (10)$$

where N is an overall noise figure. If the limit of detectability is reached when the output voltage is equal to the noise output voltage,

$$\Delta V_s = G(ZNkT\Delta f)^{1/2} \quad . \quad . \quad . \quad . \quad (11)$$

or, from eqn. (7),

$$\frac{(\Delta P)^2}{4P} = NkT\Delta f \quad . \quad . \quad . \quad . \quad . \quad (12)$$

This is to be compared with the expression

$$\delta P = NkT\Delta f \; . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

for the sensitivity of a radar receiver.

The power, P, available at the end of a waveguide absorption cell of length l is related to the power, P_0, introduced into the guide by

$$P = P_0 \exp -(\alpha_w + \alpha_g)l \quad . \quad . \quad . \quad . \quad (14)$$

where the attenuation coefficient, α_w, describes the losses in the waveguide walls and α_g is the absorption coefficient of the gas for particular conditions of pressure and temperature. The change in power produced by the gas is thus given by

$$\begin{aligned} \Delta P &= P_0 \exp -\alpha_w l - P_0 \exp -(\alpha_w + \alpha_g)l \\ &= P_0 \exp -\alpha_w l(1 - \exp -\alpha_g l) \\ &= P_0 \alpha_g l \exp -\alpha_w l \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15) \end{aligned}$$

We may now obtain an expression for the minimum detectable gaseous absorption coefficient by combining eqns. (12) and (15):

$$P_0 \alpha_g l \varepsilon^{-\alpha_w l} = \sqrt{(4P_0 \varepsilon^{-\alpha_w l} NkT\Delta f)}$$

or

$$\alpha_g = \frac{2}{l}\sqrt{\frac{NkT\Delta f}{P_0 \varepsilon^{-\alpha_w l}}}$$

There is clearly an optimum length l which minimizes α_g. This is given by $l = 2/\alpha_w$, and thus the minimum detectable absorption coefficient is given by

$$(\alpha_g)_{min} = \varepsilon\alpha_\omega \sqrt{\frac{NkT\Delta f}{P_0}} \quad . \quad . \quad . \quad . \quad (16)$$

Similar expressions have been obtained for spectrometers using various modulations described later, which approximate to linear detection. For instruments employing crystal detectors in the square-law region the expression involves P_0^{-1} instead of $P_0^{-1/2}$.

Sensitivities of the best experimental spectrometers fail to achieve the ideal given in eqn. (16) by factors of 30 or more, indicating that N is of the order of 1 000. It might appear from eqn. (16) that the sensitivity could be improved indefinitely by increasing the source power P_0. However, for powers approaching a milliwatt, crystal noise greatly exceeds amplifier noise and N becomes dependent largely on the crystal detector* and proves to be power dependent. Crystal noise power increases with power level[7] according to the approximate relationship

$$P_n = \left(\frac{\beta P^2}{f} + 1\right) kT\Delta f \quad . \quad . \quad . \quad . \quad (17)$$

where f is the intermediate or modulation frequency and β is a parameter which varies somewhat from crystal to crystal. On the other hand, at these power levels the conversion loss L decreases with increasing power, i.e.

$$L = \frac{l}{SP} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

Understandably, it is found experimentally that, so long as the power level is sufficient for crystal noise to be dominant, the sensitivity or signal/noise ratio is practically independent of power level over a wide range. Below a few tenths of a microwatt, crystal noise becomes less important and increasing conversion loss leads to falling sensitivity. Above a few tenths of a microwatt, conversion loss no longer improves appreciably with power level. Using bandwidths of 1 c/s or less, sensitivities corresponding to $\alpha \rightarrow 10^{-10}\,\mathrm{cm}^{-1}$ have been reached.

(3) DESIGN OF SPECTROMETERS

The simplest type of spectrometer is shown in Fig. 8. Microwave power is transmitted through a waveguide absorption cell

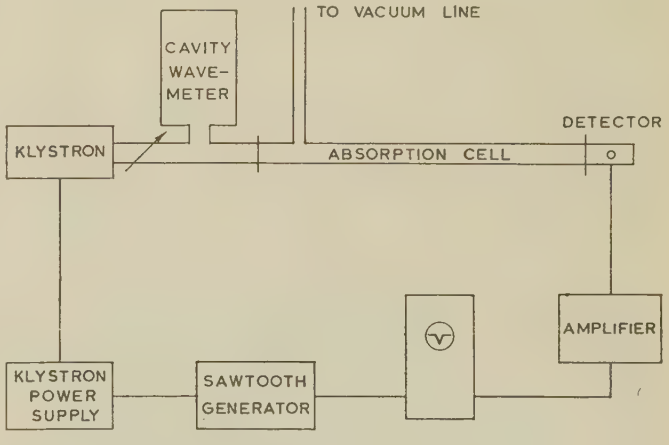


Fig. 8.—Simple microwave spectrometer.

containing the gas under investigation. The signal detected by a crystal is displayed on an oscillograph after amplification. The frequency of the source is swept through a range of about 5–20 Mc/s by a repetitive low-frequency sawtooth modulation, which is also applied to the X-plates of the oscillograph. An absorption line appears as a sharp dip in the mode curve of the microwave source.

* N, in fact, varies between crystals, and crystal selection is an important factor in obtaining maximum sensitivity.

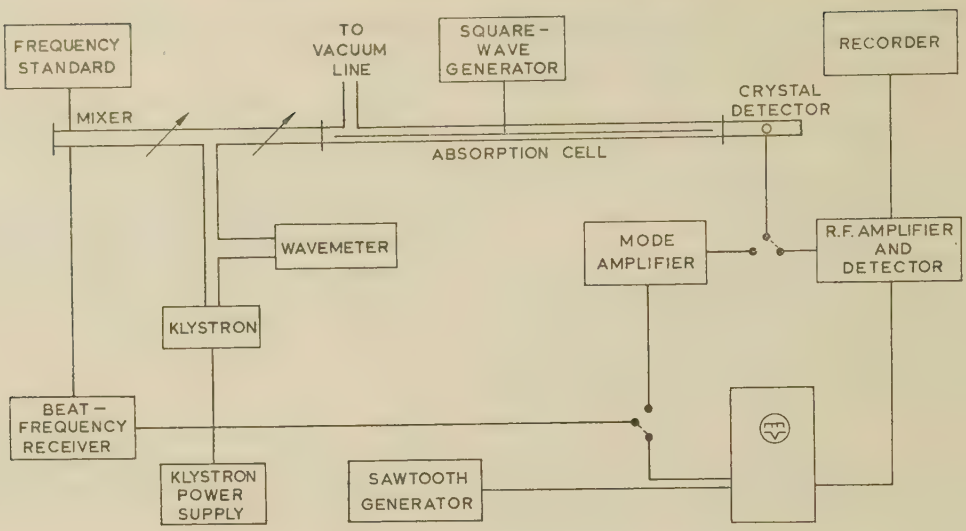


Fig. 9.—Stark modulated spectrometer.

Klystrons are the most commonly used sources; the power requirements are entirely adequate and sawtooth frequency modulation of about 20 Mc/s is readily obtained. The absorption cell, generally fitted with mica windows, may be as long as 50 ft, and is often coiled for convenience in cooling in Drikold or liquid air, to improve absorption intensities.

This type of instrument is simple to construct and operate. Its sensitivity is limited to detecting lines with absorption coefficients greater than about $10^{-6}\,cm^{-1}$. Reduction in bandwidth is achieved only by using lower sweep frequencies, and this is offset by an increase in crystal noise according to eqn. (17). Irregularities in the mode curve and reflections in the waveguide absorption cell present serious difficulties in searching for weak lines.

(3.1) Molecular Modulation

Of the various modulation methods which have been employed, Stark modulated instruments[8] are the most widely used general-purpose instruments.

(3.1.1) Stark Modulation.

In instruments of this kind[9, 10] the gaseous sample is subjected to an approximately uniform electric field which is generally applied as a periodic zero-based square-wave, although sine-wave modulation has sometimes been employed. The operation of such an instrument may be understood by reference to Fig. 4. If the source frequency is f_1, microwave power will be absorbed during field-free half-cycles. When the source frequency is f_2 or f_3, there will be absorption during the half-cycles when the field is applied but not during the field-free half-cycles. Thus for a frequency absorbed by the gas, the microwave power becomes amplitude modulated at the frequency of the square-wave Stark voltage; for frequencies which are not absorbed such amplitude modulation is absent. The detected signal is amplified by a narrow-band tuned amplifier. The method has the very important advantage of largely overcoming the otherwise troublesome effects of reflections and klystron mode irregularities which become limiting factors in searching for weak lines with unmodulated instruments. In addition, provided that sufficient power is available, a gain in sensitivity is achieved over the simple spectrometer because of the reduction in noise power per unit bandwidth according to eqn. (17).

The main features of a Stark modulated instrument are shown in Fig. 9. For maximum sensitivity the detected signal is fed into a pre-amplifier matched to the impedance of the crystal appropriate to the microwave power employed. This is often followed with a narrow-band communication receiver, but for maximum sensitivity it is necessary to reduce noise still further and to this end bandwidths of about 1 c/s are commonly attained by the use of an amplifier frequently described as a lock-in amplifier, phase-sensitive detector or homodyne amplifier. Such amplifiers have the additional advantage of providing a detected signal in which field-free lines f_1 and Stark lines f_2 and f_3 are distinguished by opposite signs. The use of such narrow bandwidths demands a slow frequency sweep to avoid distortion of line shape, and the spectrum is then normally traced out by a pen-recorder.

The choice of square-wave modulation frequency is governed by two factors. Consideration of crystal noise which depends on frequency according to eqn. (17) favours a high modulation frequency. On the other hand, increase of modulation frequency leads to broadening of absorption lines by producing sidebands. To a good approximation the width of the line is increased to

$$\Delta f_m = \Delta f_0\left[1 + \frac{l}{4}\left(\frac{f}{\Delta f_0}\right)^2\right] \quad . \quad . \quad . \quad . \quad (18)$$

where Δf_0 is the width of the line in the absence of modulation and f is the modulation frequency.[11] This sets an upper limit of about 100 kc/s on the modulation frequency. Frequencies as low as a few kilocycles per second have been used with advantage where sharp lines are required for high-resolution work.

The design of the square-wave generator[13] is determined largely by the required voltages and properties of special absorption cells employed.[9, 12] Probably the most widely used type of cell is that shown in Fig. 10(*a*). It is conveniently made of

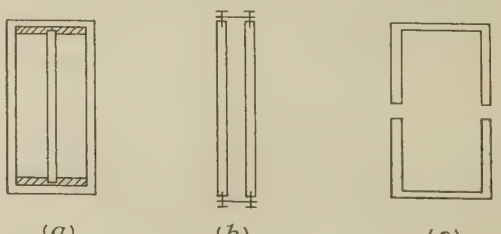


Fig. 10.—Waveguide absorption cells for Stark modulated instruments.

(*a*) Rectangular waveguide with central electrode supported by Teflon tape.
(*b*) Parallel plates separated by mica spacers.
(*c*) Rectangular waveguide cut along the no-current line.

out 3 m of K- or X-band waveguide, through the entire length which there passes a central electrode perpendicular to the lines. The ends of the electrode are tapered to reduce reflections. Teflon tape provides good insulation from the waveguide, chemically inert and has a low loss. Parallel plates insulated mica spacers at intervals [Fig. 10(*b*)] have been used to provide waveguide cells and Stark electrodes at the same time. These ells have advantages at high frequencies, where the first type produce large attenuation. Fig. 10(*c*) shows a simple type of cell readily made by cutting a waveguide along the no-current line. It suffers from poor homogeneity of the Stark field.

Cells of the type (*a*) and (*b*) have capacitances of about 500 pF per a length of 3 m. Sine-wave modulation is readily made with a low-power oscillator using resonant conditions. Square-wave modulation, which is spectroscopically advantageous, requires the much higher power, CV^2f, and if the slopes of the wavefront and wavetail are not to exceed 1% of the pulse length, the applied voltage must provide for peak currents of $200\,CVf$. Square-wave generators providing these peak currents and voltages of 1–1·5 kV normally operate by charging and discharging the Stark cell by a triggered pair of thyratrons or—for higher frequencies—pulse beam tetrodes.

(3.2) **Zeeman Modulation.**

The majority of molecules in their ground state do not possess magnetic moments. There are, however, a small number of molecules, e.g. O_2, NO and ClO_2, which are paramagnetic in their ground states and many unstable molecules, e.g. OH and S, and excited molecules have magnetic dipole moments. For these cases there arises the possibility of magnetic-field (Zeeman) modulation. Considerable effort has been put into the development of Zeeman modulated spectrometers as a means for the detection and study of free radicals.[14] Such instruments assemble the Stark modulated instrument shown in Fig. 10, except for the cell and the associated modulator. A waveguide cell is placed in a solenoid which provides sine-wave magnetic modulation. The cell is enclosed in a glass tube to contain the gas and is frequently sectioned, as shown in Fig. 10(*c*), to avoid eddy currents. Up to the present time the spectra of only a small number of free radicals have been observed, the main difficulty being to build up a sufficient concentration of molecules of short life in a waveguide cell.

(3.2) Source Modulation

In a source-modulated instrument the klystron is frequency modulated over a small range compared with the spectral line width. At the same time the frequency is swept through the line by low-frequency sawtooth modulation or by mechanical tuning. The absorption line behaves as a discriminator, converting the frequency-modulated signal to an amplitude-modulated one. The signal is amplified and displayed as in a Stark modulated spectrometer. Source modulation is experimentally considerably simpler than Stark modulation, but it suffers from the disadvantage that changes in the mode power and reflections in the cell appear in the detected signal and frequently have an appearance resembling spectral lines. In practice these effects become limiting in searching for weak lines. This factor, together with the additional spectroscopic information obtained by Stark modulation, account for the rather restricted use of source modulation.

(3.3) Frequency Measurements

Sufficiently accurate measurements for such purposes as line identification may be made with cavity wavemeters. For the evaluation of sufficiently accurate moments of inertia to yield precise molecular structures, a frequency precision of 0·05–0·01 Mc/s is generally required. In a typical frequency standard which achieves this accuracy a series of multipliers carries a frequency through the sequence 5, 20, 60, 180 and 540 Mc/s. The 5 Mc/s oscillator, controlled by a thermostat-controlled crystal, is monitored against the MSF standard 5 Mc/s signal to better than 1 c/s. The 540 Mc/s signal is further multiplied and mixed with the unknown microwave frequency and the difference frequency is determined. The final frequency multiplication and mixing are conveniently made on a single crystal. Q-band crystals are found to be very effective for this purpose and have been used to measure microwave frequencies up to 40 Gc/s operating on about the 75th harmonic of the input frequency.[15] A conventional crystal holder is used, the crystal is separated from the waveguide mount by a thin paper insulator, the 540 Mc/s signal is fed into one of the crystal connectors and the difference frequency is taken from the other.

(4) SPECIAL-PURPOSE SPECTROMETERS

Microwave spectroscopy is limited to the study of polar molecules in the gas phase at low pressure. Increasing the pressure leads to broader lines without increase of peak intensity, and at lower pressures the lines become too weak to observe. Two problems arise—extending the technique to deal with substances not sufficiently volatile at room temperatures and reducing line width for high-resolution work.

(4.1) High-Temperature Spectrometers

For temperatures up to about 150° C conventional instruments can be modified. Teflon spacers for the Stark electrode are satisfactory and lead gaskets can be used for the mica windows. The temperature range can be extended to about 300° C by replacing the Teflon spacer by mica supports. The extension of the temperature range to 1 000° C proved to be a problem of a different order of magnitude, for the oxidation of copper or brass waveguides at these temperatures presents a serious problem. A cell made from stainless-steel waveguide fitted with a Stark electrode and enclosed in a Vycor envelope has been used.[16] Another very successful solution has been to use a nickel waveguide with a gold-plated inner surface.[17] Displacement and distortion of the Stark electrode during heating are minimized by mounting the cell vertically. Ceramic spacers are used, and double vacuum-spaced windows avoid any appreciable pressure gradient across the hot quartz window. Connections to the Stark electrode and the heating coil are made from tantalum. With this spectrometer the spectra of such substances as the alkali halides as gaseous diatomic molecules have been obtained.

(4.2) High-Resolution Instruments

The so-called 'natural' line width of an absorption line is determined by the time a molecule spends in the rotational state before losing energy as spontaneous radiation. Such a width is appropriate to a system of isolated molecules at rest. For a low-pressure gas several factors lead to an increase in the width many orders of magnitude greater than the extremely narrow natural line width of microwave absorption of about 10^{-4} c/s. At pressures above about 10^{-2} mm Hg, pressure broadening, brought about by perturbation of the rotation by collisions, is dominant. At lower pressures (about 10^{-3}–10^{-4} mm Hg) collisions with walls become comparable in importance with inter-molecular collisions. In addition, there is the pressure-independent Doppler width arising from the distribution of molecular velocities relative to the direction of microwave propagation. Experimental methods may introduce the further effects of Stark modulation broadening and saturation broadening, which increases with increasing microwave power. High-

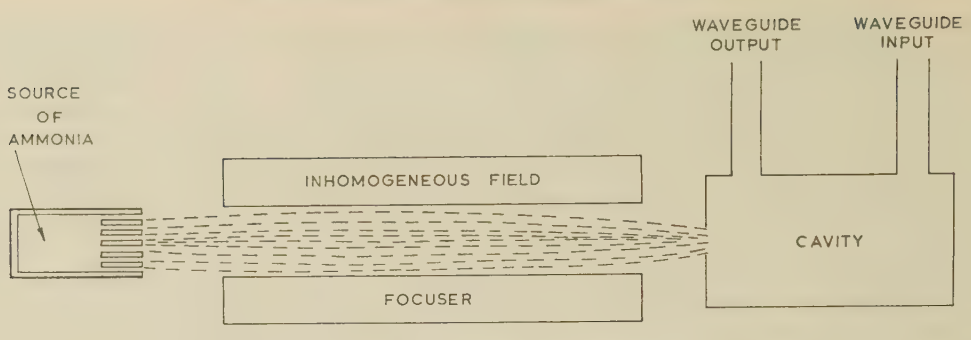


Fig. 11.—The gaseous maser.

resolution spectroscopy thus requires the use of low pressures (10^{-3}–10^{-4} mm Hg), low microwave power (a few microwatts) and, if Stark modulation is used, a low modulation frequency. X-band waveguide cells are preferable to smaller sizes in helping to avoid broadening due to wall collisions and saturation. Stark modulated instruments operated at a few kilocycles per second have been employed, but these instruments lose some of their normal advantage in sensitivity through the large conversion loss at the low powers required. Instruments using superheterodyne detection (described in Section 4.4) avoid this difficulty.

A balanced-bridge instrument using an absorption cell in conjunction with a dummy cell and superheterodyne detection[18] has recorded lines with a width of 60 kc/s. By the incorporation of 1 kc/s Stark modulation into this instrument, the limitation on sensitivity imposed by reflections, etc., in a non-modulated instrument has been reduced without loss of resolution. In an instrument of this kind Doppler broadening is a major source of line width. The use of low temperatures reduces this effect, but only in a limited way; the Doppler effect falls off as $T^{-1/2}$, but the exponential decrease of vapour pressure with the temperature often results in volatility of the sample limiting the possible reduction in temperature. Line widths below the Doppler widths have been obtained by observing the absorption of a plane wave passing through a beam of ammonia molecules moving perpendicular to the wavefront.[19] The main component of molecular velocity under these conditions is parallel to the wavefront, and a line width of 20 kc/s was obtained instead of the normal Doppler width of 36 kc/s for this line. A microwave absorption line width of only 350 c/s has been achieved for caesium using an atomic-beam technique.[20] Such narrow line widths are important in the design of frequency standards.

(4.3) The Gaseous Maser

The use of molecular-beam and microwave spectroscopic methods have been elegantly combined in the development of the first maser.[21] With this instrument molecular absorption line widths of about 7 kc/s were achieved.

The construction of the maser is shown diagrammatically in Fig. 11. Ammonia gas streams out of the source vessel into a vacuum sufficiently low that a beam is formed and passes into the electrostatic focuser. Here four parallel electrodes which run along the length of the focuser are shaped and carry voltages in such a manner as to produce an inhomogeneous electric field which has its minimum along the axis and increases radially. Consider the effect on ammonia molecules in the two inversion states represented in Fig. 6. The two are seen to behave in quite different ways in an electric field. The Stark effect leads to an increase in energy for the molecules in the upper state, and such molecules experience a force towards the axis of the focuser, while molecules in the lower state move outwards along a radius to regions of higher fields. Thus the beam entering the cavity consists mainly of molecules in the upper state.

It follows from the discussion of induced emission (Section 1.4) that, if radiation of the appropriate frequency is fed into the cavity, induced emission at that frequency will occur. There results the 'microwave amplification by stimulated emission of radiation', from which the device takes its name. By stimulating the cavity over a frequency range, the spectrum may be explored. Since the molecular velocities are essentially along the axis of the cavity, line widths several times smaller than the Doppler widths are obtained. In this way a new hyperfine structure was revealed. Besides its spectroscopic interest, the maser is also of interest because of its possibilities as a microwave amplifier, oscillator and frequency standard.

(4.4) Superheterodyne Spectrometers

From the importance of superheterodyne receivers in radar it might appear that such receivers would be important in spectroscopy. They have been built, but their use has been restricted mainly to special-purpose work. Normally in spectroscopy the signal is not a weak pulse but a small decrement in a comparatively large amount of power. The consequence is that only when the spectroscopic source power is low do superheterodyne receivers have advantages over Stark modulated instruments, for which the sensitivity may be limited by conversion loss at low powers. Even this is valid only if a local oscillator is available with moderate power, and in the region where source power is low because of the need to use harmonic generators, this is frequently not so. Such instruments have the disadvantage that, in searching for a spectrum, two klystrons—the source and the local oscillator—must be tuned simultaneously, keeping a constant intermediate frequency. This proves to be too cumbersome for general use.

(5) APPLICATION TO CHEMICAL ANALYSIS

(5.1) Qualitative Analysis

The high resolution available in microwave spectroscopy is an attractive feature for identification of chemical substances. Almost always the frequency of a single line serves to identify a compound, and certainly the measurement of two lines places identification beyond doubt. There are no characteristic frequencies of molecular groups, as in infra-red spectra. On the other hand, the spectra of similar molecules are very easily distinguished, often having little or no resemblance. Even the very similar molecules $^{14}N^{15}N^{16}O$ and $^{15}N^{14}N^{16}O$ have absorption frequencies as widely separated as 25 121·56 and 24 274·60 Mc/s for the transitions between the 0 and 1 quantum levels.

For identification purposes the spectrometer need not be complicated, and sufficiently accurate frequency measurements can be made with a cavity wavemeter. Spectra can be obtained from small recoverable samples; microgram quantities fill a normal absorption cell. The sensitivity of the method is good, but it is difficult to indicate the limits precisely, since these

pend on a number of factors, especially molecular dipole oment and absorption frequency. In the case of normal rbon oxysulphide, lines are easily seen for the isotopic species ^{13}CS in natural abundance of 1%.

Up to the present time the main experimental difficulty in perating a microwave spectrometer as an automatic recording strument has resided in the source. For the range approximately from 10 cm to 7 mm, klystrons have normally been used, nd from about 7 mm to 1 mm, crystal harmonic generators are enerally used. Several klystrons are required if it is necessary cover the complete range, and in some regions these need to e selected, or even fitted with modified cavities. In some frequency regions it is possible to sweep through about 1 Gc/s by avity tuning. On the other hand, in regions falling between the entre frequencies of two standard klystrons the effective sweep nge may be reduced by the frequent need to optimize conditions r power and mode shape. Where harmonic generators must e used the need for continuous adjustments during searching is ven greater. While these effects present difficulties for the utomatic recording of spectra over a large frequency range, ey are no serious barriers to recording a spectrum over a elected range of 1 Gc/s or so. This would often be sufficient identify an asymmetric rotor molecule. The development of ackward-wave oscillators may further make possible the autoatic recording of spectra over wider frequency ranges.

(5.2) Quantitative Analysis

Quantitative analysis presents the rather more difficult problems of measuring and interpreting intensities.[22] Absolute bsorption intensities are not easily measured, but relative ntensities may be satisfactorily measured for analytical purposes. t has already been seen that for a pure substance the peak ntensity increases with increasing pressure up to a limiting value. Further increase of pressure simply changes the width of the line y collision broadening. In this second pressure region, ncreasing the pressure of a foreign gas also leads to line roadening and at the same time to a fall in peak intensity, eaving the integrated intensity under the absorption line unchanged. Under certain conditions the peak intensity gives a measure of the fractional abundance of the absorbing species. The method is particularly attractive for measuring abundances of isotopic molecular species and has been applied in this way. Wider application of the method requires calibration for each foreign substance in a mixture, since each substance has a specific line-broadening power. The use of integrated intensities would avoid the need for calibration, but further experimental work is required before such intensities may be obtained sufficiently accurately.

The main limitation on the experimental accuracy of measuring relative peak intensities of lines arises from variations in importance of reflections within the waveguide cell at the different line frequencies. Recent developments in this field include a Stark modulated cavity cell[23] and the use of 'antimodulation'.[24]

(6) REFERENCES

(1) Gordy, W., and Sheridan, J.: 'Microwave Spectra and Structures of Methyl Mercuric Chloride and Bromide', *Journal of Chemical Physics*, 1954, **22**, p. 92.

(2) Millen, D. J., and Morton, J. R.: 'The Microwave Spectrum of Nitric Acid', *Journal of the Chemical Society*, 1960, p. 1523.
Morton, J. R.: 'The Microwave Spectrum and Structure of Nitric Acid', Ph.D. Thesis, University of London, 1957.

(3) Mizushima, M.: 'On the Ammonia Molecule', *Physical Review*, 1948, **74**, p. 705.

(4) Townes, C. H., and Geschwind, S.: 'Limiting Sensitivity of a Microwave Spectrometer', *Journal of Applied Physics*, 1948, **19**, p. 795.

(5) Gordy, W.: 'Microwave Spectroscopy', *Reviews of Modern Physics*, 1948, **20**, p. 668.

(6) Strandberg, M. W. P., Johnson, H. R., and Eschbach, J. R.: 'Apparatus for Microwave Spectroscopy', *Review of Scientific Instruments*, 1954, **25**, p. 776.

(7) Torrey, H. C., and Whitmer, C. A.: 'Crystal Rectifiers', Radiation Laboratory Series, Vol. 15 (McGraw-Hill, 1948).
Nicoll, G. R.: 'Noise in Silicon Microwave Diodes', *Proceedings I.E.E.*, Paper No. 1671 R, September, 1954 (**101**, Part III, p. 317).

(8) Hughes, R. H., and Wilson, E. B.: 'A Microwave Spectrograph', *Physical Review*, 1947, **71**, p. 562.

(9) McAfee, K. B., Hughes, R. H., and Wilson, E. B.: 'A Stark Effect Microwave Spectrograph of High Sensitivity, *Review of Scientific Instruments*, 1949, **20**, p. 821.

(10) Sharbaugh, A. H.: 'The Design and Construction of a Stark-modulated Microwave Spectrograph', *ibid.*, 1950, **21**, p. 120.

(11) Karplus, R.: 'Frequency Modulation in Microwave Spectroscopy', *Physical Review*, 1948, **73**, p. 1027.

(12) Baird, D. H., Fristrom, R. M., and Sirvetz, M. H.: 'Stark Effect Absorption Cells for Microwave Spectroscopy', *Review of Scientific Instruments*, 1950, **21**, p. 881.

(13) Hedrick, L. C.: 'A Flexible High-Voltage Square-Wave Generator', *ibid.*, 1949, **20**, p. 781.

(14) Sanders, T. M., Schawlow, A. L., Dousmanis, G. C., and Townes, C. H.: 'A Microwave Spectrum of the Free OH Radical', *Physical Review*, 1953, **89**, p. 1158.

(15) Jackson, R.: 'The Microwave Spectrum and Structure of Chlorine Monoxide', Ph.D. Thesis, University of London, 1959.

(16) Tate, P. A., and Strandberg, M. W. P.: 'A Simple High-Temperature Microwave Spectrograph', *Review of Scientific Instruments*, 1954, **25**, p. 956.

(17) Stitch, M. L., Honig, A., and Townes, C. H.: 'A High-Temperature Microwave Spectrometer', *ibid.*, p. 759.

(18) Geschwind, S.: 'High Resolution Microwave Spectroscopy', *Annals of the New York Academy of Sciences*, 1952, **55**, p. 751.

(19) Johnson, H. R., and Strandberg, M. W. P.: 'Beam System for Reduction of Doppler Broadening of Microwave Absorption Lines', *Physical Review*, 1952, **85**, p. 503.
Strandberg, M. W. P., and Dreicer, H.: 'Doppler Line Width Reduction', *ibid.*, 1954, **94**, p. 1393.

(20) Essen, L., and Parry, J. V. L.: 'The Caesium Resonator as a Standard of Frequency and Time', *Philosophical Transactions of the Royal Society*, 1957, **250***A*, p. 45.

(21) Gordon, J. P., Zeiger, H. J., and Townes, C. H.: 'Molecular Microwave Oscillator and New Hyperfine Structure in the Microwave Spectrum of Ammonia', *Physical Review*, 1954, **95**, p. 282, and, 'The Maser—New Type Microwave Amplifier Frequency Standard and Spectrometer', *ibid.*, 1955, **99**, p. 1264.

(22) Hughes, R. H.: 'Chemical Analysis with a Microwave Spectrograph', *Annals of the New York Academy of Sciences*, 1952, **55**, p. 872.

(23) Verdier, P. H., and Wilson, E. B.: 'Relative Intensities of Microwave Absorption Lines', *Journal of Chemical Physics*, 1958, **29**, p. 340.

(24) Dymanus, A.: 'Intensity Measurements in Microwave Spectroscopy: the "Antimodulation Method" ', *Physica*, 1959, **25**, p. 859.

621.372.832.43

The Institution of Electrical Engineers
Paper No. 3350 E
Jan. 1961

RECENT ADVANCES IN THE USE OF COUPLED TRANSMISSION LINES AS DIRECTIONAL COUPLERS

By RYUTARO KOIKE, B.Sc.(Eng.), Graduate.

(*The paper was first received* 1*st April, and in revised form* 8*th August*, 1960.)

SUMMARY

If two similar coaxial transmission lines are coupled by sharing a common outer conductor for part of their length, so as to form a 3-conductor line, they constitute a symmetrical directional coupler.

In work on coaxial directional couplers at microwave frequencies up to 7 Gc/s, new designs of coupled transmission lines have been developed for use as power monitors and variable attenuators.

LIST OF SYMBOLS

a, a_1, a_2 = Radius of inner conductor or round conductor.
b = Radius of outer conductor.
c = Radius of inner conductor.
d, d_1, d_2, g = Geometrical distance between conductors of the transmission lines.
h, h_1, h_2 = Length used for expressing a slot size.
i_1, i_2 = Instantaneous current.
K = Constant.
k = Coupling coefficient, $L_{12}/\sqrt{(L_{11}L_{22})} = C_{12}/\sqrt{(C_{11}C_{22})}$.
L_{11}, L_{22} = Self-inductance of line 1 or 2 per unit length.
L_{12} = Mutual inductance between lines 1 and 2 per unit length.
l = Length of the coupled transmission line.
p, s, s' = Geometrical distance between conductors of the coupled transmission line.
t = Time.
V_F, V_G = Output voltage.
V_i = Input voltage.
v_1, v_2 = Instantaneous voltage.
β = Phase-change coefficient.
θ, ψ, ϕ = Angle subtended by the slots at the axis of inner conductors, rad.

(1) INTRODUCTION

The use of directional couplers as a means of measuring reflections for monitoring power and for mixing has been known for many years. Oliver[1] has given a reasonably simple and necessarily general treatment of the directional coupling action of ordinary transmission lines. He showed that the natural coupling between parallel transmission lines is inherently directional. A similar phenomenon was investigated many years ago in trying to eliminate crosstalk in telephone transmission systems.[2] Monteath[3] pointed out some features in the behaviour of two similar unbalanced transmission lines which are coupled by sharing a common outer conductor. He made the directional couplers from two lead-covered polythene-filled cables (Uniradio 25) and carried out investigations at frequencies between 180 and 600 Mc/s.

The paper describes theoretical and experimental work carried out on coupled coaxial lines which are used (*a*) as a symmetrical directional coupler for frequencies up to 7 Gc/s, (*b*) as a power monitor, and (*c*) as a variable attenuator. With regard to (*b*), since the power-carrying cable is, in general, large compared with the cable carrying the monitored signal, coupling between lines of unequal size is considered. Application (*c*) utilizes variable coupling between cables.

COUPLED TRANSMISSION LINES IN GENERAL

The following assumptions will be made throughout the paper:

(*a*) The losses are negligible, i.e. the propagation coefficient is a purely imaginary number.
(*b*) The conductor sizes and spacings are small compared with the wavelength.
(*c*) All currents are confined to the surfaces of the conductors.
(*d*) The dielectric medium is homogeneous and isotropic.
(*e*) Only TEM waves propagate on the lines.

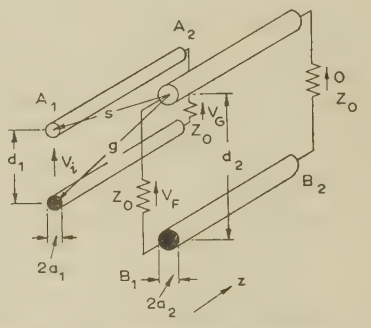


Fig. 1.—Coupled transmission lines.

The propagation of the waves on two parallel transmission lines such as those shown in Fig. 1 is described* by the transmission-line equations[1,2]

$$\left.\begin{aligned}
\frac{\partial v_1}{\partial z} + L_{11}(1 \pm k)\frac{\partial i_1}{\partial t} &= 0 \\
\frac{\partial v_2}{\partial z} + L_{22}(1 \pm k)\frac{\partial i_2}{\partial t} &= 0 \\
\frac{\partial i_1}{\partial z} + C_{11}(1 \mp k)\frac{\partial v_1}{\partial t} &= 0 \\
\frac{\partial i_2}{\partial z} + C_{22}(1 \mp k)\frac{\partial v_2}{\partial t} &= 0
\end{aligned}\right\} \quad \ldots\ldots (1)$$

The coupling coefficient, k, is given by

$$k = \frac{\log g/s}{\sqrt{(\log d_1/a_1 \log d_2/a_2)}} \quad \ldots\ldots (2)$$

for two parallel transmission lines, as shown[1] in Fig. 1. If two

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Mr. Koike, of the Anritsu Electric Co., Ltd., Japan, is at present in the Department of Electrical Engineering, University College, London.

* We are concerned with the normal modes of the system. In this case we have four normal modes (two in each direction) characterized by fixed ratios of v_2/v_1 and i_2/i_1 which do not change as the wave propagates. These ratios are characterized by $v_2/v_1 = \pm\sqrt{(L_{22}/L_{11})} = \pm\sqrt{(C_{11}/C_{22})}$ and $v_1 i_1 = v_2 i_2$.

nsmission lines are similar, i.e. if $a_1 = a_2 = a$ and $d_1 = d$, eqn. (2) becomes

$$k = \frac{\log g/s}{\log d/a} = \frac{\log \sqrt{[1 + (d/s)^2]}}{\log d/a} \quad . \quad . \quad . \quad (3)$$

iver[1] derived the output voltages on each line of the two rallel lines, coupled over a length l and each terminated in its aracteristic impedance Z_0. The output voltage on each line is en by

$$V_G = \frac{\sqrt{(1-k^2)}}{\sqrt{(1-k^2)}\cos \beta l + j \sin \beta l} V_i \text{ at } A_2 \quad . \quad . \quad (4)$$

$$V_F = jk \frac{\sin \beta l}{\sqrt{(1-k^2)}\cos \beta l + j \sin \beta l} V_i \text{ at } B_2 \quad . \quad (5)$$

ence the coupling factor is given by

$$\frac{V_F}{V_G} = j \frac{k}{\sqrt{(1-k^2)}} \sin \beta l \quad . \quad . \quad . \quad . \quad (6)$$

hich varies sinusoidally with length and frequency. When the ngth of the coupling region is a quarter wavelength, the upling factor becomes a maximum, $k/\sqrt{(1-k^2)}$. It is then tually independent of frequency, since the coupler covers a : 1 band between the points at which the coupling factor ops by 3 dB. However, as the length is reduced the coupling comes increasingly frequency-dependent until for a very small upling hole the coupling factor changes by 6 dB per octave. his latter characteristic is identical with that of the Bethe hole upler.

For small coupling ($d/s \ll 1$), the numerator of eqn. (3), g $\sqrt{[1 + (d/s)^2]}$, tends to $\frac{1}{2}(d/s)^2$ and k becomes infinitesimal, . $k/\sqrt{(1-k^2)} \simeq k$. Consequently the maximum coupling ctor, in decibels, is proportional to $\log s$.

Therefore the coupling factor in decibels is approximately

$$20 \log_{10} k = K - 40 \log_{10} s \quad . \quad . \quad . \quad . \quad (7)$$

here K is a constant.

(3) EXPERIMENTAL WORK

(3.1) Symmetrical Directional Coupler

Fig. 2 shows the directional coupler. Measurements of upling factor and directivity were carried out on couplers

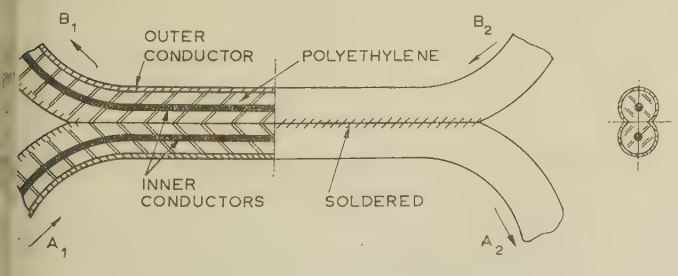


Fig. 2.—Coaxial directional coupler.

nade from 50-ohm Japanese standard cable of the following limensions:

(*a*) Inner conductor 1·4 mm; outer conductor 5 mm.
(*b*) Inner conductor 2·2 mm; outer conductor 8 mm.

Two copper-covered polyethylene-filled cables were each nilled to form a flat surface parallel to the longitudinal axis of he cable $\lambda/4$ long, so that the dielectric was exposed in a slot in he outer conductor. The cables were then clamped together with the slots coinciding, and the outer conductors were soldered together. The slots were necessarily tapered at each end, so that the actual slot length was longer than $\lambda/4$. The slot length was determined by the following method.

As shown in Fig. 3, h_1 and h_2 are the distances between the

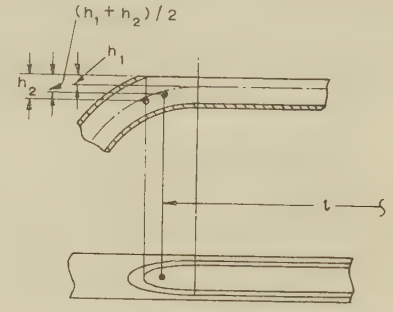


Fig. 3.—Determination of the effective slot length.

surface of the slot and the centre of the inner conductor, at the uniform coupling section and at the end of the slot, respectively. We can find the points where the distance from the surface to the inner conductor is $(h_1 + h_2)/2$ at both ends of the slot. The effective length of the slot is assumed to be the separation of these two points. A number of couplers were made and examined at various frequencies. The results showed that this assumption is quite reasonable. When TEM waves propagate through polyethylene media, the wavelength becomes about 67% of that in free space. The coupling coefficient k of this coupled transmission line can be obtained by the method of Appendix 8.1 and eqn. (3).

(3.2) Coupler for use as a Power Monitor

Very high levels of microwave power, as used in scatter propagation, are employed. For very-high-power measurement, the coupling action between two different-sized coaxial lines has been studied. The primary line which carries high power must be large in size, while the secondary line is not necessarily large since it handles power 20–40 dB below that of the primary line. The dimensions of the two cables used for the experiment are as stated in the previous Section. The width and length of each slot were made equal. Measurements were made as in the previous Section. The coupling coefficient is given as before by the method of Appendix 8.1 and eqn. (2).

(3.3) Variable-Coupling Directional Coupler

Lack of a practical variable coaxial attenuator often creates considerable difficulties for the experimental worker. When a directional coupler is used as a microwave mixer an attenuator is required to adjust the input signal level. If the coupling is variable the coupler can be used for adjusting the signal level.

Fig. 4 shows one of the variable-coupling directional couplers

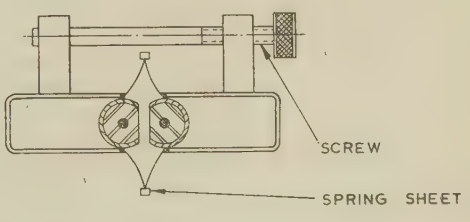


Fig. 4.—Cross-section of variable-coupling directional coupler.

which were constructed. Two copper-covered polyethylene-filled cables were milled to form a flat surface, as described in the previous Section. They were then clamped together by metal springs which also served to screen the conductors. The relation between the coupling and the spacing of the two inner conductors can easily be derived; this device is a type of reactance attenuator in which the pick-up loops are of substantial length compared with the wavelength. The analysis given in Appendix 8.2 shows that the coupling factor in decibels is proportional to the logarithm of the spacing of the two inner conductors.

(4) EXPERIMENTAL RESULTS

(4.1) Couplers of Fixed Coupling[4]

The characteristics of symmetrical and non-symmetrical directional couplers made from 50-ohm polyethylene cable are illustrated in Fig. 5, which shows the coupling as a function

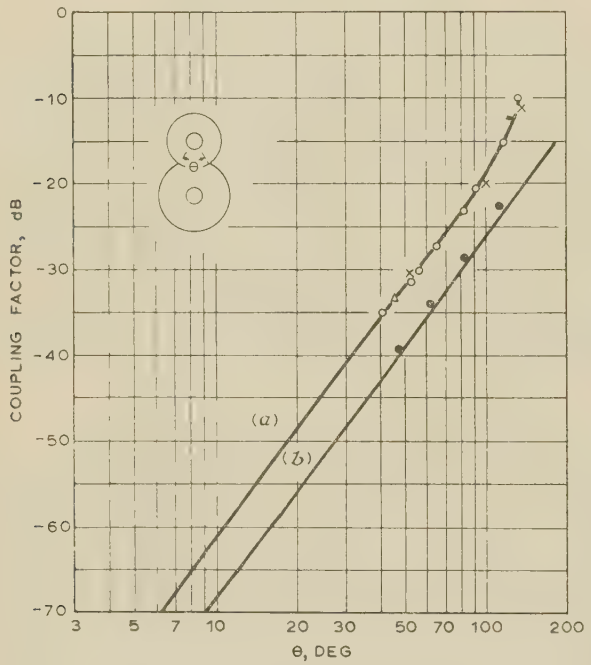


Fig. 5.—Characteristics of the directional couplers of fixed-coupling type.

(*a*) Symmetrical type.
(*b*) Non-symmetrical type.
——— Theoretical maximum coupling.
△ Measured coupling at 400 Mc/s band.
□ Measured coupling at 1 300 Mc/s band.
○ ● Measured coupling at 2 000 Mc/s band.
× Measured coupling at 4 000 Mc/s band.
■ Measured coupling at 6 000 Mc/s band.

of θ, the angle subtended by the slot at the axis of the smaller inner conductor. The technique used in the measurement of the directivity was the same as in Reference 5. The coupling factor and the directivity were measured by the precision cut-off attenuator of the receiver. The results for the symmetrical coupler are shown in curve (*a*), and those for the asymmetrical coupler in curve (*b*). A number of directional couplers of the symmetrical type with coupling slots of different size have been made for use at frequencies between 400 and 7 000 Mc/s. The non-symmetrical couplers were examined in the 2 Gc/s band. The maximum coupling, i.e. when the slot length is a quarter wavelength, is plotted against θ in Fig. 5. The experimental curves show good agreement with the theoretical ones. The characteristics of some couplers, whose slot lengths are such as to obtain maximum coupling at 2·2 and 4 Gc/s, showed coupling within ±0·2 dB for frequencies between 1·9 and 2·4 Gc/s and between 3·6 and 4·3 Gc/s, respectively. The directivity of these couplers varies with frequency. It exceeds 25 dB at frequencies lower than 3 Gc/s, but falls to 21 dB in the 4 Gc/s band. It was found that a relationship exists between the slot shape and directivity. However, at present, it is not completely known.

(4.2) Couplers of Variable Coupling

Fig. 6 shows the experimental data for the variable directional coupler. At the design frequency of 2·2 Gc/s, and with the

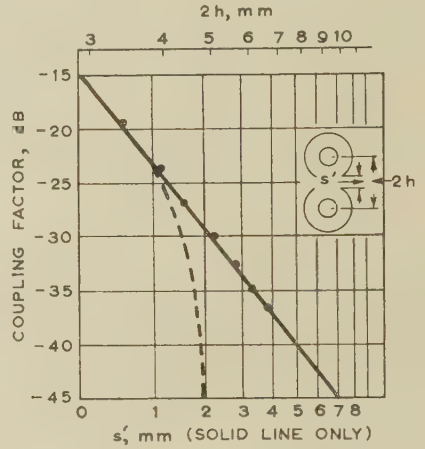


Fig. 6.—Characteristics of the variable directional coupler.

——— Logarithmic linear relation described in Section 8.2.
— — — — Calculated maximum coupling based on the theory of Section 8.1.
● Measured maximum coupling.

cables in actual contact, i.e. the condition of maximum coupling, the coupling factor was −15 dB. The analytical results are also shown in Fig. 6; the solid line shows the results of the analysis in Section 8.2, while the dotted curve shows the coupling of the fixed coupler in the previous Section as a function of the distance between two inner conductors, $2h$. The analysis given in Section 8.2 shows that the coupling is closer than that of the dotted line. The experimental data satisfy a logarithmic linear law to within the limits of experimental error. Fig. 7 shows the

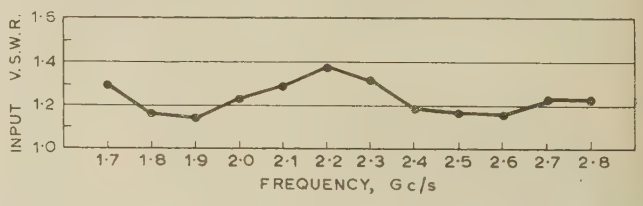


Fig. 7.—Input v.s.w.r. of the variable-coupling directional coupler. The separation is 3 mm.

input v.s.w.r. of the coupler when the separation is 3 mm. The v.s.w.r. is less than 1·2 when two lines are clamped together. It increases with larger spacing, which is an undesirable feature. The directivity also becomes less with larger spacing, but it exceeds 15 dB in all cases. For signal maximuming a high directivity is not essential, 10 dB being adequate.

(5) CONCLUSION

Close agreement has been obtained between experiment and theory for directional couplers using coupled coaxial lines. The theory of coupled transmission lines is shown to be adequate for the design of a microwave directional coupler. The couplers

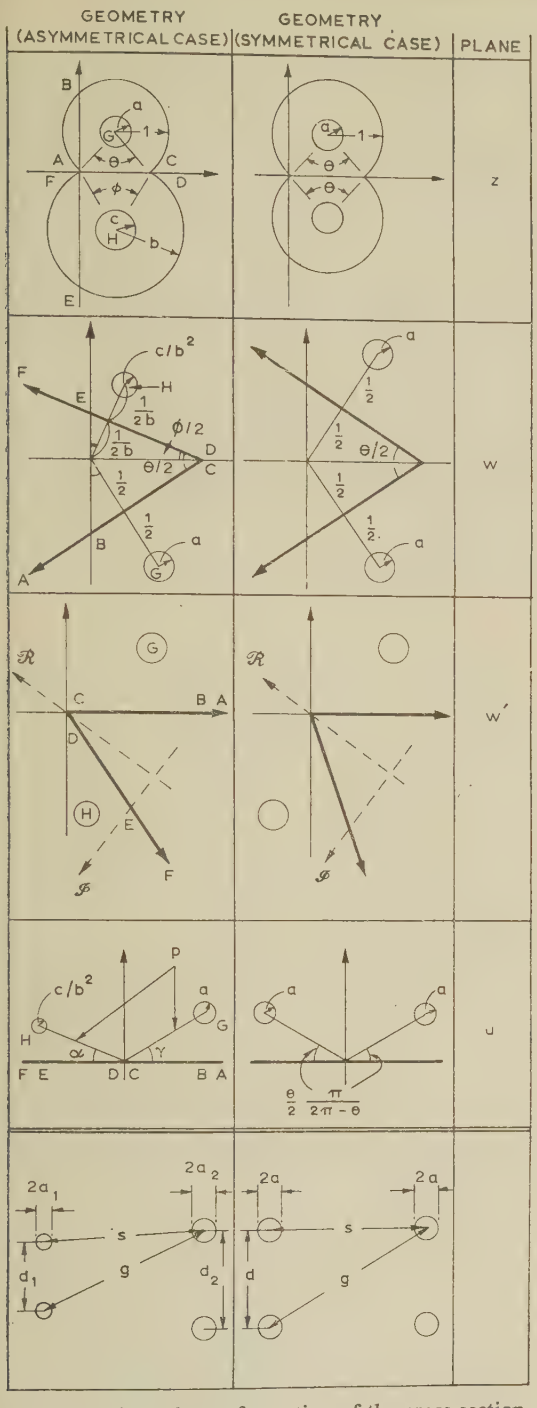


Fig. 8.—Conformal transformation of the cross-section.

Transformations z to w: $w = 1/z$.

w to w': Move the origin to the apex of the wedge and rotate $(\pi - \theta/2)$ in an anti-clockwise direction.

$$w' \text{ to } u: \frac{du}{dw'} = (2w' \sin \theta/2)^{\frac{\pi}{2\pi - (\theta + \phi)/2} - 1}$$

$$\text{or } u = \frac{2\pi - (\theta + \phi)/2}{2\pi \sin \theta/2}(2w' \sin \theta/2)^{\frac{\pi}{2\pi - (\theta + \phi)/2}}$$

$$= p(2w' \sin \theta/2)^{\frac{\pi}{2\pi - (\theta + \phi)/2}}$$

$$a_1 = c/b^2, \quad a_2 = a$$

$$d_1 = 2p \sin \alpha, \quad d_2 = 2p \sin \gamma$$

$$s^2 = (d_2 - d_1)^2/4 + p^2(\cos \alpha + \cos \gamma)^2$$

$$g^2 = (d_2 + d_1)^2/4 + p^2(\cos \alpha + \cos \gamma)^2$$

$$\alpha = \frac{\phi}{2} \frac{\pi}{2\pi - (\theta + \phi)/2}$$

$$\gamma = \frac{\theta}{2} \frac{\pi}{2\pi - (\theta + \phi)/2}$$

described have been used for microwave power measurement, power monitoring and signal mixing.

(6) ACKNOWLEDGMENTS

The paper is published by permission of the Director of the Anritsu Electric Co., Ltd. The author wishes to thank his colleagues, Messrs. A. Kishimoto, T. Asano, A. Morishima and M. Shirotori for their helpful suggestions.

The author also wishes to express his appreciation of the kind help given by Prof. H. E. M. Barlow, Dr. J. Brown, and Mr. G. D. Monteath.

(7) REFERENCES

(1) OLIVER, B. M.: 'Directional Electromagnetic Couplers', *Proceedings of the Institute of Radio Engineers*, 1954, **42**, p. 1686.

(2) CARSON, J. R., and HOYT, R. S.: 'Propagation of Periodic Wave over a System of Parallel Wires', *Bell System Technical Journal*, 1927, **4**, p. 495.

(3) MONTEATH, G. D.: 'Coupled Transmission Lines as Symmetrical Directional Couplers', *Proceedings I.E.E.*, Paper No. 1833 R, May, 1955 (**102** B, p. 383).

(4) KOIKE, R., MORISHIMA, A., and ASANO, T.: 'Coaxial Directional Couplers', *Journal of the Institute of Electrical Communication Engineers of Japan*, 1958, **41**, p. 626.

(5) GINZTON, E. L., and GOODWIN, P. S.: 'A Note on Coaxial Bethe-Hole Directional Couplers', *Proceedings of the Institute of Radio Engineers*, 1950, **38**, p. 308.

(8) APPENDICES

(8.1) Analysis of the Coupling between two Different-Sized Coaxial Lines[4]

It is required to find the parallel-wire transmission-line equivalent of the 3-conductor line. Because the conductors are parallel the associated electric field is the same in each plane perpendicular to the axis of conductors. Accordingly, the problem of determining the electric field is essentially a 2-dimensional one. The method is to transform the cross-section of the 3-conductor line into a simpler form by conformal representation. It is essentially identical with Monteath's analysis,[3] but a more general case is described here. The transformation for the non-symmetrical case is illustrated in Fig. 8, together with the symmetrical case. For convenience the radius of the smaller outer conductor is taken to be unity, and the width of the slot is expressed in terms of θ (or ϕ), the angle subtended at the axis of inner conductor of the smaller (or larger) line. The angles θ and ϕ are related by $\sin \theta/2 = b \sin \phi/2$; b is the radius

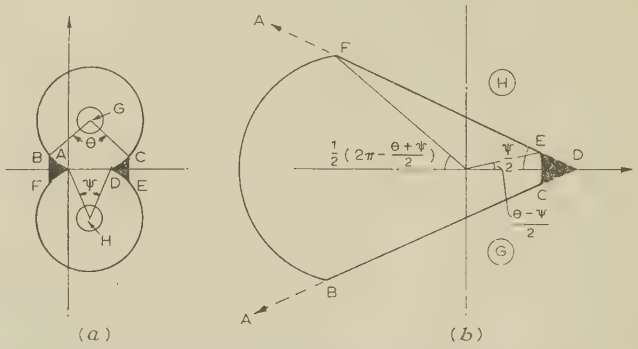


Fig. 9.—Conformal transformation of the cross-section.

(*a*) z-plane.
(*b*) w-plane.

of the larger outer conductor and is not less than unity. It should be noted that, by the first transformation $w = 1/z$, the radius of the smaller conductor is unchanged, while the other is changed from c to c/b^2. From Fig. 8 the coupling coefficient k is given by eqn. (2) or (3).

(8.2) Analysis of the Variable-Coupling Directional Coupler

The coupling for the symmetrical coaxial coupler is to be determined when the two lines are slightly separated. Although an exact solution is not obtained, the following considerations give an adequate indication of the results.

As indicated in Fig. 9(*a*), the cross-points A and D are found by extending the periphery of two outer conductors. For convenience the width of the slot is expressed as before in terms of θ, and the spacing of two slotted lines is expressed in terms of ψ, the angle $\angle$AGD. Taking one of the cross-points A as the origin of the z-plane and carrying out an inverse transformation $w = 1/z$, as in the previous Section, the z-plane is transformed into the w-plane [Fig. 9(*b*)]. Straight lines CE and BF in the z-plane are converted into the arcs of circles, CE and BF respectively, in the w-plane. The effect of the areas ABF and CDE in the z-plane will be understood from the diagram in the w-plane. If the spacing of two coaxial lines tends to increase in Fig. 9(*a*), the apex of the wedge moves to the left. Thus the coupling between the two inner conductors becomes closer to that of the case analysed in the previous Section. When the two lines are separated so that no cross-point is found, i.e. loose coupling, the logarithmic linear law [eqn. (7)] is applicable. The denominator of eqn. (3) is the impedance of the parallel transmission line itself. It is therefore analogous to the separation of two slotted coaxial lines, and thus the geometrical constant k is as in eqn. (3), and the coupling factor is written as in eqn. (7).

The dotted line in Fig. 6 shows the coupling with the edges ABF and CDE as shown in Fig. 9(*a*) in the z-plane, while the solid line indicates the coupling without these edges. This solid line satisfies the linear relation between the coupling in decibels and $\log s'$.

621.375.9

The Institution of Electrical Engineers
Paper No. 3342 E
Jan. 1961

AN ELECTRON-BEAM PARAMETRIC AMPLIFIER FOR THE 200 Mc/s REGION

By G. O. CHALK, M.Sc.

(*The paper was first received* 15*th January, and in revised form* 13*th July,* 1960.)

SUMMARY

The paper describes the design and performance of a low-noise electron-beam parametric amplifier of the quadrupole type. This is a transverse-field device in that the electron beam interacts with a radio-frequency field transverse to its direction of motion.

Particular attention is directed to the conditions necessary to achieve the lowest overall noise figure of the amplifier. It is shown that these also lead to the broadest amplification bandwidth.

The minimum noise figure so far obtained is 1·6 dB at about 200 Mc/s, which corresponds to an amplifier noise temperature of 170°K. This was obtained at an overall gain of 10 dB which required about 2 mW of pump power. The amplification bandwidth was 4 Mc/s. These performance figures compare well with those reported for the original quadrupole amplifier at higher frequencies.

LIST OF SYMBOLS

B = Flux density of axial magnetic field, Wb/m^2.
B_b = Flux density of Brillouin value of magnetic field, Wb/m^2.
C = Capacitance of input and output coupler plates, F.
F_a = Amplifier noise figure.
F_m = Measured noise figure of amplifier plus receiver.
F_r = Receiver noise figure.
G_b = Beam loading conductance at ω_{s0}, mhos.
G_c = Shunt conductance of the tuned circuits due to losses, mhos.
G_g = Internal conductance of signal source, mhos.
G_L = Load conductance terminating output coupler, mhos.
I = Electron-beam current, amp.
$\mathscr{P}$ = Electron beam perveance, amp/volt$^{3/2}$.
L = Inductance of input and output coils, H.
$L(\omega_s)$ = Unpumped transmission loss at ω_s, dB.
M_a = Overall amplifier gain, dB.
M_e = Electronic gain in quadrupole at $\omega_{s0}(=\omega_c)$, dB.
M_e' = Electronic gain in quadrupole at $\omega_s(\neq\omega_c)$, dB.
N = Number of electron orbits in a coupler.
P_L = Power loss in a tuned circuit due to coil losses, watts.
P_{max} = Saturation power output, watts.
Q_L = Loaded Q-factor of a tuned circuit.
V = Electron-beam voltage, volts.
V_p = Peak value of pump voltage, volts.
Y_b = Beam loading admittance, mhos.
a = Electron-beam diameter, m.
a_0 = Electron-beam diameter for Brillouin flow, m.
d = Coupler plate spacing, m.
l = Coupler plate length, m.
l_q = Axial length of quadrupole, m.
n = Turns ratio of the tappings on the input and output coils.
s = Quadrupole radius, m.
t_c = Transit time of an electron in a coupler, sec.
t_q = Electron transit time in quadrupole, sec.
u_0 = Axial electron velocity, m/s.
Δ = Fractional amplification bandwidth.
η = Charge/mass ratio for an electron, C/kg.
$\theta = (\omega_c - \omega_s)t_c$, rad.
ρ = Electron orbital radius at exit to quadrupole, m.
ρ_0 = Electron orbital radius at entrance to quadrupole, m.
ω_c = Cyclotron frequency of electrons, rad/s.
ω_p = Pump frequency, rad/s.
ω_s = Signal frequency, rad/s.
ω_{s0} = Centre frequency of signal band, rad/s.

(1) INTRODUCTION

Many experimental results have been published on parametric amplifiers of the variable-capacitance semiconductor-diode type, but little has been reported on the performance obtained with the various electron-beam amplifiers, the only one so far reported to have achieved low-noise performance being the quadrupole amplifier devised by Adler.[1,2] Although somewhat more complex than the variable-capacitance diode amplifier, the electron-beam parametric amplifier has a number of practical advantages, particularly in its behaviour under large-input-signal conditions.

When subjected to a large input signal such as, for example, transmitter breakthrough in a radar equipment, the semiconductor diode may easily suffer permanent damage owing to overheating of the very small junction. On the other hand, the electron-beam amplifier will normally either saturate, as in the longitudinal-field devices such as the travelling-wave tube, or have zero output because the beam is defocused, as in the transverse-field devices. In both devices normal operation is restored when the signal overload is removed and, in addition, a well-designed valve should be free from large-signal paralysis. These may be very important considerations in a practical application and outweigh the apparent simplicity of the semiconductor amplifier.

The transverse-field or quadrupole amplifier is one of a number of electron-beam parametric amplifiers which have been proposed, but it is the only one which has so far proved successful as a low-noise device. The other valves which have been suggested are longitudinal-field devices, and although parametric gain has been obtained, their noise figure has been high, typically about 20 dB. The quadrupole amplifier described by Adler, on the other hand, has achieved a noise figure as low as 1·3 dB. A further advantage of the quadrupole amplifier is its extreme stability. The only connection between the input and output being via the beam, the tube is unilateral in its action and therefore unconditionally stable. Most of the variable-capacitance amplifiers are of the negative-resistance type and as such are regenerative, so that input and output isolators are essential in order to obtain stability at high levels of gain. The amplification bandwidth of the negative-resistance amplifier is inversely proportional to the voltage gain, so that at high levels of gain the bandwidth is low, voltage-gain/fractional-bandwidth products of 10^{-2} being typical. In the quadrupole amplifier the bandwidth is determined solely by the input and output couplers which impress and remove the signal from the beam, while the gain is determined solely by the design of the quadrupole section. For this reason the bandwidth is independent of the gain and

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Mr. Chalk is with the English Electric Valve Co. Ltd.

voltage-gain/fractional-bandwidth products of about unity have been obtained.

(2) THE QUADRUPOLE AMPLIFIER

The quadrupole amplifier, named after the quadrupole configuration of electrodes used to produce the parametric gain in the device, is shown diagrammatically in Fig. 1. A cylindrical

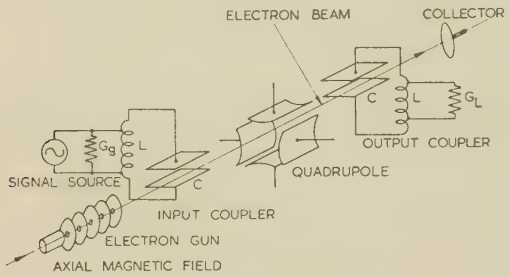


Fig. 1.—The quadrupole amplifier.
The quadrupole connections are shown in Fig. 2. For identical couplers, $G_L = G_g$.

electron beam is focused by an axial magnetic field of such a value that the cyclotron frequency of the electrons is equal to the signal frequency. This beam passes between three sets of radio-frequency electrodes to a collector. The first electode system, the signal-input coupler, consists of a pair of deflector plates of capacitance C. These are tuned to the signal frequency by a coil of inductance L coupled, via a tapping, to the signal-input source of conductance G_g. This is an electron coupler of the type described by Cuccia[3] in which r.f. energy from a signal source is converted into rotational kinetic energy of the electrons forming the beam. The input signal power produces an alternating electric field between the coupler plates transverse to the direction of motion of the beam, causing electrons normally travelling in straight lines parallel to the axis to describe helical paths about the tube axis with radii increasing linearly in time as they pass through the coupler. In this way the signal energy is converted into rotational energy of the orbiting electrons. At the exit from the coupler the radius of the electron orbit is directly proportional to the amplitude of the signal field across the coupler plates.

Without pump power applied to the quadrupole, the electrons leaving the input coupler continue to spiral at a constant radius until they enter the output coupler, which for simplicity is assumed to be identical to the input one. (This is not essential but simplifies the discussion. In the actual valve identical couplers are employed.) Provided that the output coupler is terminated in a load conductance, G_L, of the correct value, the orbiting electrons give up all their rotational energy to the load as they pass through the coupler. In doing so the radii of their orbits decrease linearly with time in the coupler, so that on leaving it their orbital radii and therefore their rotational energies are zero.

The same process removes the transverse or rotational noise motion from the beam in the input coupler when it is terminated in a conductance of the correct value. The transverse components of noise in the beam entering the input coupler from the electron gun is absorbed in the terminating conductance as the beam passes through the coupler, leaving it in a virtually noise-free condition in the transverse sense. Since this particular value of the terminating conductance is identical to that which gives a maximum power transfer from the signal-input source on to the beam, the coupler performs the dual role of removing the noise and impressing the signal on to the beam simultaneously.

Amplification of the electron orbital motion between the two couplers is brought about by the conversion of energy from the r.f. pump to the electrons, using the quadrupole structure. This is tuned by four coils to the pump frequency, which is twice the cyclotron frequency. The quadrupole is shown in cross-section in Fig. 2.

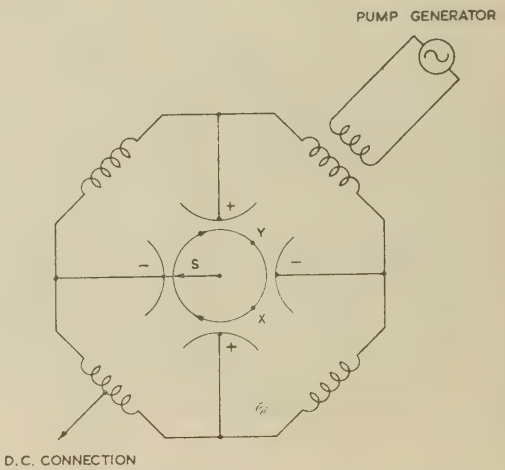


Fig. 2.—Cross-sectional view of the quadrupole.
The signs indicate the instantaneous polarity of the electrode voltages. The straps connecting opposite plates to ensure π-mode operation are omitted.

The quadrupole electrode system produces an r.f. electric field with radial and azimuthal components rotating about the axis at one-half the pump frequency, i.e. the cyclotron frequency. An orbiting electron which, for example, is instantaneously at point X will always be in an accelerating field during its passage through the quadrupole and so will gain energy from the field, spiralling outwards to an orbit of larger radius. On the other hand, an electron at point Y will be in a retarding field and so will give up energy to the field, in doing so spiralling inwards to an orbit of smaller radius. However, the effect on all electrons entering the quadrupole averaged over all phases of the pump field is a net gain, the total energy extracted from the pump field exceeding the total energy given up to it. The magnitude of the quadrupole field at any point is directly proportional to the radial distance from the axis, so that an electron with a larger orbital radius initially, corresponding to a stronger input signal, will experience a stronger quadrupole field. This is the condition necessary for exponential gain. The gain mechanism does not depend on the signal frequency and is constant for a fixed pumping-power input, the amplification bandwidth being determined solely by the selectivity of input and output couplers.

As in all parametric amplifiers, if the signal frequency is not equal to half the pump frequency (in this case the cyclotron frequency) an 'idler' of frequency equal to the difference between the pump and signal frequencies will be generated.

(3) VALVE DESIGN

In designing this valve it was intended to aim for a performance comparable with the original quadrupole amplifier, namely a power gain of 20 dB, a fractional bandwidth of 10% and a noise figure of about 2 dB.

It was stated in Section 1 that the couplers and the quadrupole perform essentially independent functions in the operation of the valve, and in the design they may therefore be considered separately, at least to a first approximation. The Cuccia couplers, of which the input one simultaneously impresses the signal and removes the noise from the beam and the output one

oves the amplified signal from the beam, determine the
o wing valve parameters:

) Noise figure.
) Unpumped transmission loss.
) Bandwidth.
) Saturation output and dynamic range.

ne quadrupole provides the gain of the valve by converting
p power at twice the cyclotron frequency into the signal
idler channels but affects the other parameters only indirectly;
example, in order to obtain high gain the pumping field
cause some defocusing and interception of the beam, result-
in a degradation of the noise performance of the valve.
he main restriction on the design of the couplers and
drupole is imposed by the electron beam which can be
ised by the available magnetic field. For this reason the
n-focusing problem will be considered first.

(3.1) The Electron Beam

n this valve the magnitude of the axial focusing magnetic field
xed by the requirement that the cyclotron frequency shall be
al to the centre frequency of the signal band. This is
n by

$$\omega_c = \omega_{s0} = \eta B \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

ch may be simplified to

$$B = 3{\cdot}6 \times 10^{-11} f_{s0} \quad . \quad . \quad . \quad . \quad . \quad (2)$$

centre frequency of 200 Mc/s implies a focusing field of
$\times 10^{-3}$ Wb/m^2. This invariance of B places limitations on
beam which can be focused.

n the quadrupole amplifier as in the low-noise travelling-wave
e it is essential to keep current intercepted by the r.f. structure
the absolute minimum in order to obtain the lowest possible
ise figure. The beam, on passing through the correctly matched
ut coupler, is rendered virtually noiseless in the transverse
se by an adjustment of the noise rotational motions of the
lividual electrons, so that the overall transverse noise of the
gregate of electrons is zero. Any electrons subsequently lost
m the beam will upset this balance and some beam noise will
refore return.

In order to obtain a beam whose profile will be least affected
any perturbing influence, such as r.f. lens effects between the
ds of the couplers and the quadrupole, it is desirable to employ
immersed-flow method of beam focusing in which lines of
agnetic flux thread both the beam and the cathode. Although
profile of a beam focused in this way varies periodically
ng its length between a minimum diameter, a_0, and a maxi-
m diameter, a, it is considerably more resistant to perturbing
lds than a beam produced from a magnetically shielded
thode, which, for a unique value of the magnetic field called
e Brillouin field, may be focused theoretically into a perfect
linder. To consider the behaviour of the beam under
mersed flow it is convenient to relate the actual magnetic
ld employed, B, to the Brillouin value for the same beam.
e Brillouin value of magnetic field, B_b, for a beam of diameter
, voltage V and perveance $\mathscr{P}$ is given by[3]

$$B_b^2 = 2{\cdot}76 \times 10^{-6} \mathscr{P} V a_0^{-2} \quad . \quad . \quad . \quad . \quad (3)$$

The maximum diameter of the immersed flow beam, a, is
lated to the magnetic field, B, and tends to a_0 as B approaches
finity. The variation of a/a_0 with B/B_b derived from results
blished by King[3] is shown in Fig. 3. The value of B, the
cusing field which may be employed, is fixed by the frequency
operation of the valve. The value of B_b for a particular
lue of beam perturbation, a/a_0, can be obtained from this

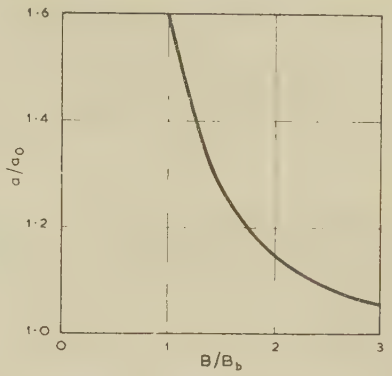


Fig. 3.—Amplitude of beam perturbation in immersed flow as a function of relative magnetic field.

value of B using the curve in Fig. 3. With eqn. (3) for Brillouin flow a range of values of a_0, $\mathscr{P}$ and V can be deduced. In this way permissible values of the beam parameters may be obtained from the specified value of the magnetic field.

Before the beam enters the input coupler it is desirable to restrict the range of transverse velocities of the electrons in the beam as much as possible, to make the complete noise-cancellation process easier to achieve. This may be accomplished by means of a collimator in the electron gun, the simplest arrangement being a pair of spaced apertured discs. In this way the outer portion of the beam may be removed, allowing a well-defined and ordered beam to enter the input coupler.

Measures must also be taken to minimize secondary emission from the collimator and the collector. This may be done by electroplating these electrodes with suitable materials and by careful geometrical design.

(3.2) The Quadrupole

A cross-sectional diagram of the quadrupole viewed along the axis of the beam is shown in Fig. 2. Although the shape of the electrodes should ideally be hyperbolic to obtain the correct field distribution, plates of circular cross-section have been used in practice with satisfactory results. The radius of curvature of the circular plates was made equal to the radius of curvature of the hyperbola at its apex.

When the signal frequency is equal to the cyclotron frequency, an electron entering the quadrupole in the correct phase with an initial orbital radius ρ_0 gains energy from the pump and leaves the quadrupole with an increased orbital radius ρ, given by

$$\rho = \rho_0 \exp\left(\frac{V_p t_q}{s^2 B}\right) \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where s is a geometrical constant shown in Fig. 2. Since the electron rotational energy is proportional to the square of the orbital radius, the electronic power gain in the quadrupole may be written

$$M_e = 8{\cdot}68 \frac{V_p t_q}{s^2 B} \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

or

$$M_e = 8{\cdot}68 \frac{V_p l_q}{s^2 B u_0} \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

When the signal frequency is not equal to the cyclotron frequency (half the pump frequency) an 'idler' of frequency $(\omega_p - \omega_s)$ is generated. In this case the electronic gain is less than that given by eqn. (5), the signal gain being equal to

$$M_e' = 10 \log_{10} \cosh^2 \left(\frac{V_p t_q}{s^2 B}\right) \quad . \quad . \quad . \quad . \quad (7)$$

and the output at the idler frequency being equal to $\sinh^2(V_p t_q / s^2 B)$ times the signal input. In the region of high gain (when the exponents are greater than two) the outputs at the signal and idler frequencies are equal and eqn. (7) may be written as:

$$M_e' = 8{\cdot}68\frac{V_p t_q}{s^2 B} - 6 \quad . \quad . \quad . \quad . \quad . \quad (8)$$

The electronic gain is independent of the signal frequency except for the special case when the latter is exactly equal to half the pump frequency, and for a particular pumping level is therefore a constant. The overall amplifier gain, however, depends on the r.f. transmission loss through the valve in the absence of pumping (unpumped transmission loss) which is a function of the signal frequency depending on the selectivity of the input and output couplers. The overall amplifier gain is therefore given by

$$M_a = M_e' + L(\omega_s) \quad . \quad . \quad . \quad . \quad . \quad (9)$$

where $L(\omega_s)$ is the transmission loss and is negative. It is desirable to minimize the pump voltage required for a particular overall gain in order to keep the intercepted current due to pumping to a minimum. Consequently it is important to make the unpumped transmission loss as low as possible. Values better than −1 dB at the band centre are typical of a well-designed tube. In this case an electronic gain of 21 dB is required to obtain an overall gain of 20 dB at the band centre.

(3.3) Input and Output Couplers

(3.3.1) Unpumped Transmission Loss.

The input coupler may be represented by the equivalent circuit in Fig. 4. L represents the total circuit inductance, C the total

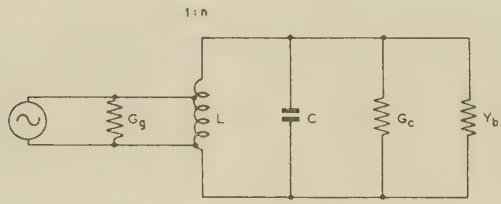


Fig. 4.—Schematic of the input coupler.

The output coupler is identical except that the signal generator is replaced by the input conductance G_L of the following stage.

circuit capacitance, n the turns ratio of the coil tap ($n > 1$), G_c the shunt conductance due to circuit losses, G_g the signal generator input conductance and Y_b the beam loading admittance. When the signal frequency is equal to the cyclotron frequency, Y_b becomes a pure conductance, G_b. The resonant frequencies of the input and output tuned circuits are made equal to the cyclotron frequency. The transmission loss is a minimum for this frequency, which is therefore adjusted to be at the band centre. The loss increases on either side, owing to the selectivity of the tuned circuits and to a susceptive component of the beam loading admittance. These will be discussed in Section 3.3.3. By proper adjustment of the coil turns ratio, the circuit and beam loading admittances can be matched to the signal generator admittance. In the ideal case of zero circuit losses, all the input signal power can be transferred to the electrons (and from the electrons to the load on the output coupler) so that the transmission loss will be zero provided that there is no loss of energy from the electron motion on passing from the input to the output coupler. In the practical case the power transfer is not complete, owing to the finite losses in the two tuned circuits. A simple calculation shows the power loss in a single tuned circuit at the band centre to be

$$P_L = 10 \log_{10} \frac{G_b}{G_c + G_b} \quad . \quad . \quad . \quad . \quad (10)$$

The overall transmission loss is due to two circuits, so that $L(\omega_s)$ at the band centre is

$$L(\omega_{s0}) = 2P_L \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

To keep a low pump level for a given overall gain it is desirable to minimize $L(\omega_s)$. This can be achieved by using a low-loss tuned circuit (G_c low) and a high beam loading conductance. For a 1 dB transmission loss at the band centre G_b must be greater than $8{\cdot}3G_c$.

(3.3.2) Noise Figure.

As explained earlier, a low noise figure will be obtained only with a well-focused electron beam with little or no interception on the structure. To cancel the transverse noise fluctuations on the beam the input coupler must be accurately matched to the input source admittance. This is also the condition for maximum power transfer from the input-signal source on to the beam so that a correct adjustment of the coil tapping ratio, n, will fulfil both requirements. However, any input circuit loss increases not only the transmission loss but also the overall tube noise figure by exactly the same amount, so that a transmission loss of, say, −0·5 dB in the input, due to coil losses, effectively adds +0·5 dB to the tube noise figure. The importance of a large ratio G_b/G_c is obvious.

To summarize, the requirements for good noise performance are:

(*a*) Accurate matching of the beam to the input source admittance.
(*b*) Low input coupler loss.
(*c*) Low beam interception both with and without pumping.
(*d*) A beam with a narrow range of low thermal transverse energies.

The fourth requirement may be interpreted by saying that the 'cooler' the beam entering the input coupler the less accurate the match need be to obtain a good noise figure. This will make the performance less sensitive to input match and a low noise figure over a wider bandwidth may be obtained.

(3.3.3) Bandwidth.

The amplification bandwidth of the tube is restricted in the first place by the superposition of the selectivity curves of the input and output tuned circuits. If both circuits have the same loaded Q-factors, the fractional bandwidth to half-gain points can be shown to be

$$\Delta = \frac{(\sqrt{2} - 1)^{1/2}}{Q_L} \quad . \quad . \quad . \quad . \quad . \quad (12)$$

If the beam loading conductance, G_b, is very much greater than the cavity shunt conductance, G_c, the loaded Q-factor of the tuned circuit in Fig. 4 is

$$Q_L = \frac{\omega_{s0} C}{G_b} \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

This also assumes that G_b does not vary with frequency. Thus, the required beam-loading conductance is

$$G_b = \frac{\omega_{s0} C \Delta}{(\sqrt{2} - 1)^{1/2}} \quad . \quad . \quad . \quad . \quad . \quad (14)$$

It is thus desirable to make the total circuit capacitance, C, as

as possible to obtain the widest bandwidth with a particular m loading conductance. At 200 Mc/s C might typically be F, so that for a 10% bandwidth G_b should be about 10^{-4} mho.

The beam loading admittance, Y_b, presented to the coupler at signal angular frequency, ω_s, may be shown to be given by[5]

$$Y_b = 2G_b\left(\frac{1-\cos\theta}{\theta^2} + j\frac{\theta-\sin\theta}{\theta^2}\right) \quad . \quad . \quad (15)$$

where

$$\theta = (\omega_c - \omega_s)t_c \quad . \quad . \quad . \quad . \quad . \quad (16)$$

is the beam loading conductance when $\omega_s = \omega_c$ and θ is therefore zero. G_b is given by[5]

$$G_b = \tfrac{1}{8}\left(\frac{l}{d}\right)^2\frac{I}{V} \quad . \quad . \quad . \quad . \quad . \quad (17)$$

The reactive component of the beam loading admittance is the result of a phase slip between the signal field and the electron orbital motion at frequencies differing from the cyclotron frequency. The result of this gradually increasing phase difference between the signal field and electron orbital motion as the electrons pass through the coupler is that electrons first gain energy from the field spiralling outwards to some maximum radius which occurs when $\theta = \pm\pi$ rad. Beyond this point the phase between the electron motion and the field is such that the electrons start to return energy to the field, resulting in a decrease of orbital radius. If the transit time and frequency are such that $\theta = \pm 2\pi$ rad, no net energy is transferred to the electrons. Eqn. (17) therefore imposes a limit on the maximum coupler transit angle which can be used to obtain any desired bandwidth. This effect of phase slip as well as the two selectivity curves must be considered when designing a coupler for any particular bandwidth. Eqn. (16) with $\theta = 2\pi$ may be interpreted another way. If N is the number of electron orbits in the coupler, the fractional 'bandwidth' between the two points of zero net-energy transfer is simply $2/N$, so that the useful bandwidth will be of the order of $1/N$.

(3.3.4) **Saturation Power Output.**

The saturation output power is determined by the output coupler. In longitudinal-field amplifiers such as travelling-wave tubes saturation is made manifest by a progressive reduction in gain as the signal input is increased above the maximum working value. Ideally this effect does not occur in transverse-field devices. The output/input relationship is linear until, owing to the increased diameter of the electron orbits, the beam current is intercepted by the r.f. structure, in this case the output coupler, and the output is zero. A sharp break in the characteristic occurs only for a beam of virtually zero thickness and in practice this is not the case, the fall in output power as the input is increased being more gradual as more of the beam is intercepted. Zero output is obtained only when total interception occurs. It is not desirable to work in the region where interception is occurring since this results in a rapid increase in overall noise figure.

The maximum power in the beam when it is just grazing the output-coupler plates is given by[3]

$$P_{max} = 0{\cdot}7 \times 10^{-12}\omega_c^2 I d^2\left(1-\frac{a}{d}\right)^2 \quad . \quad . \quad (18)$$

where a is the maximum diameter of the beam. A large saturation output and therefore a large dynamic range may be obtained by using a large beam current and coupler-plate separation and small beam diameter. The maximum current and minimum beam diameter are determined by the focusing conditions. An increase in the coupler-plate spacing causes a decrease in the beam loading conductance and a consequent reduction in bandwidth and increase in transmission loss. This can be offset by a corresponding increase in coupler-plate length to keep G_b constant, but this may place a limit on the bandwidth owing to the effect of phase slip discussed in Section 3.3.3. The saturation output is a factor which is almost completely determined by the demand for a low transmission loss and a broad bandwidth.

(3.4) Complete Valve Design

The overall design of this valve, as with the majority of electron-beam devices, is a compromise between a number of conflicting factors. The overwhelming consideration, however, is to obtain as low a noise figure as possible with sufficient gain to make it useful in front of a second stage with a noise figure in the region of 10 dB.

To obtain the lowest overall noise figure the effective input tuned-circuit loss must be kept to a minimum. This calls for a low-loss circuit and a high beam loading conductance. This requires a large value of the coupler-plate aspect ratio l/d and a high beam conductance I/V. The minimum value of d is dictated by the electron-beam focusing conditions. The maximum value of l is limited by the desirability of keeping the overall length of the tube short to save on both size and weight of the focusing electromagnet. In order to obtain high gain in the quadrupole over as short a distance as possible a low beam voltage is essential for a long transit time, t_q.

The bandwidth of the valve is determined by eqns. (14) and (16). A high beam loading conductance is again desirable. The transit time in the coupling plates and therefore the coupler length should be short but, provided that N is less than about 10, phase slip in the coupler will not be the deciding factor and the bandwidth will be mainly dependent on the circuit loaded Q-factors and therefore G_b.

A number of calculations were made using a wide range of the various parameters discussed above, the aim being to obtain the lowest noise figure with reasonable gain and making a compromise between bandwidth and saturation requirements. Identical couplers were used for the input and output and the following design parameters were fixed upon:

Beam voltage	= 6 volts
Beam current	= 40μA
Beam diameter	= 0·040 in
Coupler length	= 1·60 in
Coupler plate spacing	= 0·08 in
Quadrupole length	= 1·60 in
Quadrupole radius, s	= 0·10 in

With a total circuit capacitance of 2 pF and a circuit loss conductance of 2×10^{-5} mho, these design values lead to:

Beam loading conductance	= $3{\cdot}3 \times 10^{-4}$ mho
Fractional bandwidth	= 8·4%
Saturation output power	= 25μW
Peak quadrupole voltage for an overall gain of 20 dB	= 5·5 volts
Unpumped transmission loss at ω_{s0}	= −0·6 dB

A photograph of the completed tube designed for a band centre of about 200 Mc/s and pumped at about 400 Mc/s is shown in Fig. 5.

The electron gun consists of a number of apertured discs, the first of which selects a small portion of the current emitted from a large oxide-coated cathode. The next three electrodes form the desired shape of beam and the final two electrodes form a collimator which allows the centre core of the beam to pass into the interaction region.

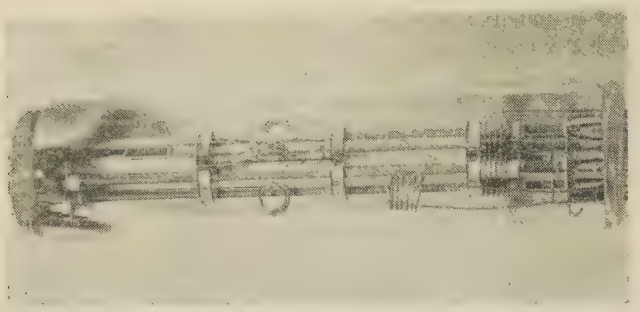


Fig. 5.—Complete valve.

The input and output couplers are tuned by equal inductances tapped to provide a correct match to the 300-ohm balanced line which passes through the valve base to the external baluns (not shown). The quadrupole is tuned to the pump frequency by four coils, the quadrupole plates being strapped to ensure operation in the π-mode. The r.f. connection is loop-coupled to one of the coils and the d.c. connection made to the centre tap of the opposite coil.

The valve is processed in a manner similar to that employed on a low-noise travelling-wave tube.

(4) PERFORMANCE

The valve was operated in an electromagnet in which the field was constant to within $\pm 2\%$ over the volume occupied by the electron beam. The solenoid was wound with copper wire with steel pole-pieces at each end. Eventually it is intended to use an aluminium-foil solenoid, when a more uniform field should be obtained.

The balanced transmission lines connecting the internal coils to the valve bases were connected via short lengths of standard 300-ohm line to balun transformers to convert the balanced outputs to standard 50-ohm coaxial line. The balun feeding the pump circuit was a simple half-wave loop and those in the input and output signal circuits were of the tuned-transformer type. These were employed so that coupler matches, in particular the input one, should be adjustable to obtain as good a match as possible to give the lowest noise figure and transmission loss.

The couplers and quadrupole were operated initially at 6 volts and the collector at 50 volts positive with respect to the cathode. The various gun voltages and the alignment of the valve in the magnetic field were adjusted until, with a total cathode current of about 5 mA, 40 μA was focused to the collector with the minimum intercepted current to the couplers and the quadrupole. Power at the centre frequency of the signal band was applied to the input circuit and the beam current and magnetic field were adjusted for minimum transmission loss. Some further small adjustment to the gun voltages was often beneficial in reducing this. Adjustment of the beam current and hence the beam loading conductance was required to take account of the slight variation in match caused by small variations in the position of the coupling coil taps. Power was then applied to the pump circuit at a frequency of 2 ω_c and gain observed. The valve was then ready for the various measurements to be performed.

(4.1) Beam Focusing

When set up as described above with a collector current of about 40 μA, the total intercepted current on the r.f. structure was typically less than 0·3 μA. This corresponds to a transmission efficiency better than 99%. When pumped to obtain an overall gain of 10 dB at the band centre (about 11 dB electronic gain) the interception remained less than 0·5 μA, but at 20 dB overall gain the interception increased to about 1 μA. This caused a degradation of the noise figure as will be shown later.

(4.2) Gain and Bandwidth

The variation with frequency of unpumped transmission loss for a typical valve is shown in Fig. 6, which is a plot of $L(\omega_s)$ of eqn. (9). At the centre of the band the loss is about 1 dB, increasing on both sides as expected from the resonant nature

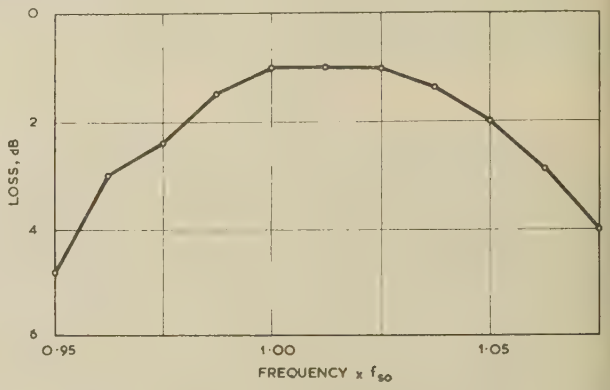


Fig. 6.—Typical variation of unpumped transmission loss with frequency.

of the tuned circuits. This loss can be accounted for almost entirely in circuit losses in the input and output coils. Cold measurements on the input and output circuits give a value of about 180 for the unloaded Q-factor of the circuits and a total circuit capacitance of about 2 pF. The calculated value of G_c is therefore $1{\cdot}5 \times 10^{-5}$ mho. Using the calculated beam loading conductance of $3{\cdot}3 \times 10^{-4}$ mho the power loss in each circuit works out at 0·2 dB. The possibility of loss of signal power from the beam between the couplers was also considered. A valve was built consisting of two Cuccia couplers only, the quadrupole section being omitted and the output coupler positioned as near as possible to the input coupler. The lowest transmission loss recorded was again about 1 dB. It can therefore be concluded that with the transit angles used in this valve the loss of signal power on the beam is negligible.

The variation of gain at the band centre with pump power is

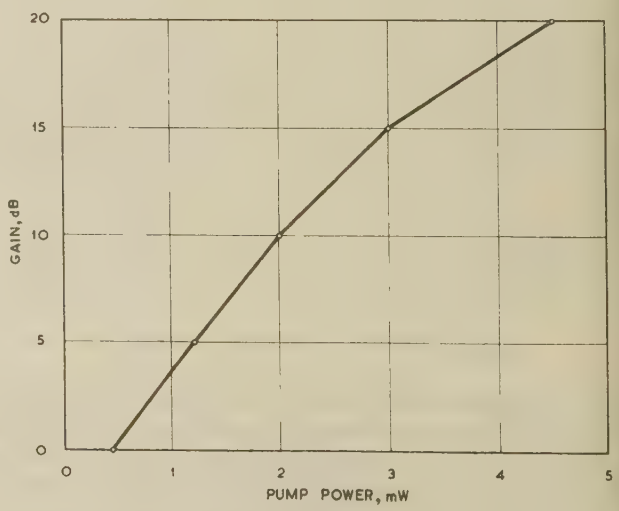
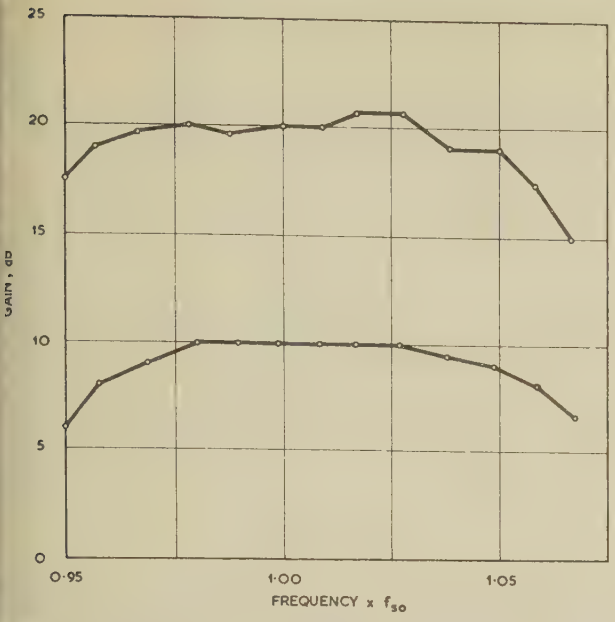


Fig. 7.—Typical variation of overall gain with pump power.



ig. 8.—Typical variations of overall gain with frequency with the pump level set for 10 dB and 20 dB mid-band gain.

own in Fig. 7. These measurements were made with the adrupole voltage reduced to 5·2 volts, where a minimum of ise occurs. Fig. 8 shows the variation in overall gain with quency, the gain being set nominally to 10 dB and 20 dB the centre of the band. The pump frequency and power d the magnetic field were unchanged during these measure- nts. This curve is very similar to the transmission-loss curve, t increases at all points by about 11 dB and 21 dB, showing the lependence of electronic gain with variation in signal frequency. The typical bandwidth of the valve to the 3 dB points is out 25 Mc/s. This is somewhat greater than the bandwidth lculated using eqn. (12), the increase being due to the effect the beam susceptance on the behaviour of the circuit at equencies away from ω_c. Since the beam susceptance is opposite sign to the susceptance of LC in parallel (Fig. 4) e total circuit susceptance will be reduced and the effective aded Q-factor also reduced. This results in an increased ndwidth of approximately the amount actually achieved.

(4.3) Noise Figure

All noise-figure measurements were made with a broad-band ise source which injected noise power into the valve at both e signal and idler frequencies, and all noise figures quoted e for the broad-band case. As in all parametric amplifiers, e noise output of the valve is the sum of three contributions, mely the internally generated valve noise, Johnson noise from e input source conductance at the signal frequency and Johnson ise from the input source conductance at the idler frequency hich is converted to the signal frequency by the parametric nplification process. Thus, when the input is a coherent signal the signal frequency, the output of the valve will consist of e amplified signal and valve noise and two contributions of hnson noise. This may be expressed in another way by ying that the input-noise bandwidth is twice the useful signal ndwidth. The additional contribution of idler noise effectively alves the input signal/noise ratio of the valve when used to ceive a signal at the signal frequency alone and adds 3 dB to e broad-band noise figure, which is measured by injecting ise power at both frequencies, so that the useful signal and noise bandwidths are equal. When the input signal is incoherent noise, as in the case of radio astronomy, the full noise capabilities of the valve can be realized.

The noise figures were measured by observing the difference in receiver output when the input termination was cooled from 290° to 90° K by immersion in liquid oxygen. This method is of use only for measuring very low noise figures when the internally generated tube noise is in the region of thermal noise. In practice two terminations are employed, one at room temperature and one in liquid oxygen, in order that they may both be good matches at their respective temperatures to ensure proper beam-noise cancellation.

Typical curves of the variation of noise figure with frequency are shown in Fig. 9. Fig. 9(*a*) was taken with the pump level

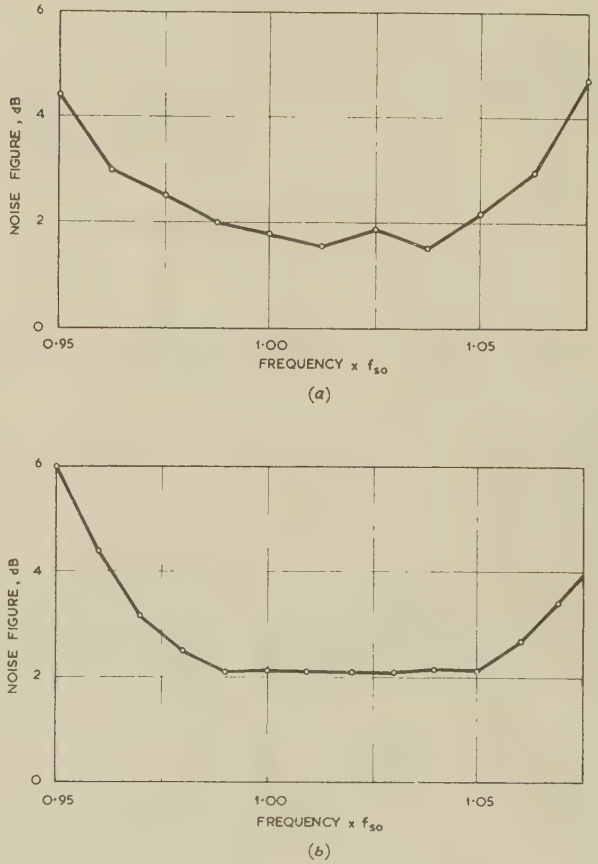


Fig. 9.—Typical variation of noise figure with frequency.
(*a*) At 10 dB mid-band gain.
(*b*) At 20 dB mid-band gain.

adjusted for a band-centre gain of 10 dB and is corrected for the effect of the noise of the second stage of the receiver using the expression:

$$F_a = F_m - \frac{F_r - 1}{M_a} \quad . \quad . \quad . \quad . \quad (19)$$

where $F_r = 6$ dB. In order to obtain the best noise figure, the various electrode voltages (in particular the quadrupole voltage) were adjusted. The heater voltage was also varied. Fig. 9(*b*) shows a curve plotted with an overall gain of 20 dB at the centre of the band. These results were taken with a different valve of the same design. In this case the gain that could be obtained

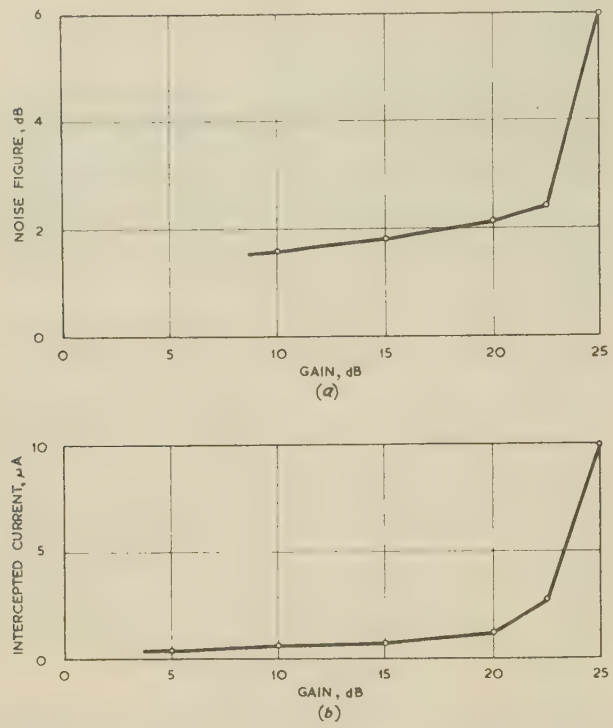


Fig. 10.—Typical variation of noise figure (*a*) and intercepted current (*b*) with gain (i.e. pump level).

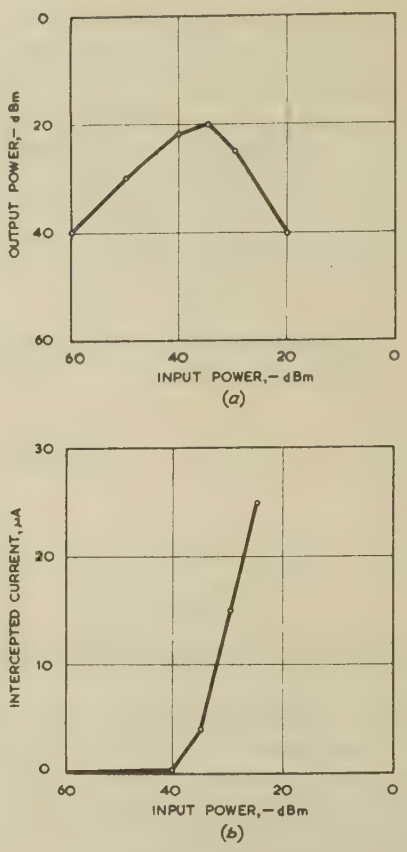


Fig. 11.—Typical variation of power output (*a*) and intercepted current (*b*) with power input.

was sufficient to make a correction for F_r unnecessary. It should be noted that the noise figure is constant over a band of some 6%. This is due almost entirely to the excellent match provided by the input balun.

The variation of noise figure with gain is shown in Fig. 10 together with the variation of intercepted current. The increase in intercepted current at higher gain levels due to the increased pump level clearly manifests itself in an increase in noise figure.

The minimum noise figure so far obtained is 1·6 dB, of which about 0·5 dB can be accounted for in input-circuit and transmission-line losses. This leaves an excess noise in the tube of about 1·1 dB. A number of reasons for this excess have been considered, the one receiving most attention at present being the design and construction of the gun. It is suspected that some, if not most, of this noise is due to incomplete noise cancellation in the input coupler, although the voltage standing-wave ratio of the input termination at the band centre viewed from the valve is better than 1·1 : 1. However, purely by adjustments to gun geometry and alignment, noise figures have been reduced from about 10 dB to the present figures, and variations in minimum noise figures of between 1·6 and 2·9 dB in identical tubes seem to be accounted for only by variations in gun alignment and cathode condition. In other words, some guns appear to produce a beam more thermally energetic than others in the transverse sense.

(4.4) Saturation

The variation of output power and intercepted current with input power at a gain of 20 dB is shown in Fig. 11. The valve behaves as a linear amplifier up to an output power of about −20 dBm when current begins to be intercepted by the structure. At this value of gain the output falls to zero at an input power of about −20 dBm. This corresponds to a dynamic range of about 70 dB.

(5) CONCLUSIONS

The results obtained from this valve show that transverse-field parametric amplifiers of the quadrupole type can be readily designed on the principles described. The performance obtained, which can be predicted with fair accuracy by simple theory, compares favourably with that of the original valves reported by Adler at higher frequencies.

(6) ACKNOWLEDGMENTS

Acknowledgments are due to Mrs. J. Sumpton and Mr. J. A. N. Snell, who carried out most of the measurements, and to Mr. C. Mulcahy for his painstaking work in constructing the valves.

The author wishes to thank Mr. J. Dain for many helpful suggestions during the preparation of the paper, and the Admiralty and the Managing Director of the English Electric Valve Co., Ltd., for permission to publish it.

(7) REFERENCES

(1) Adler, R., and Hrbek, G.: 'A Low-Noise Electron-Beam Parametric Amplifier', *Proceedings of the Institute of Radio Engineers*, 1958, **46**, p. 1756.

(2) Adler, R., Hrbek, G., and Wade, G.: 'The Quadrupole Amplifier', *ibid.*, 1959, **47**, p. 1713.

(3) Cuccia, C. L.: 'The Electron Coupler—A Developmental Tube for Amplitude Modulation and Power Control at Ultra-High Frequencies', *RCA Review*, 1949, **10**, p. 270.

(4) King, P. G. R.: 'Electron Guns for Travelling Wave Tubes', *S.E.R.L. Technical Journal*, 1954, **4**, p. 9.

(5) Smith, L. P., and Shulman, C. I.: 'Frequency Modulation and Control by Electron Beams', *Proceedings of the Institute of Radio Engineers*, 1947, **35**, p. 644.

..317.794: 621-372.822

The Institution of Electrical Engineers
Paper No. 3392 E
Jan. 1961

THE DESIGN AND PERFORMANCE OF TRANSVERSE-FILM BOLOMETERS IN RECTANGULAR WAVEGUIDES

By J. A. LANE, M.Sc., Associate Member, and D. M. EVANS, Ph.D., B.Sc., A.Inst.P.

(*The paper was received* 18*th August*, 1960.)

SUMMARY

Equations are derived which determine the design of film-type lometers used for the measurement of microwave power. The sults are discussed in relation to the concept of impedance in rect- gular waveguides.

Measurements on an improved tunable-type instrument at a wave- gth of 3 cm indicate that power levels between 1 and 100 mW can be asured in terms of a d.c. calibration with an error of not more an ±2%.

(1) INTRODUCTION

Preliminary experiments on film bolometers at microwave equencies have been described in previous papers.[1,2] These vestigations confirmed that thin metallic films, supported in e centre of rectangular waveguides on thin mica or glass, can used to measure microwave power in terms of a d.c. calibra- on. A film of the appropriate resistivity, when followed by a aveguide short-circuit at a distance slightly greater than $\lambda_g/4$, ts as a non-reflecting termination (λ_g = wavelength in the tide). The e.m.f. generated in a thermocouple attached to the m serves as a convenient indication of the absorbed power.

Further experiments on improved instruments of this type have en carried out, and the paper discusses some of the factors volved in the design of bolometers of optimum bandwidth. n addition, some results are given which indicate the accuracy btained in a tunable model intended for use in the 3 cm aveband.

(2) THEORETICAL CONSIDERATIONS

The film is located symmetrically in the transverse cross- ction of a rectangular waveguide, as shown in Fig. 1. The

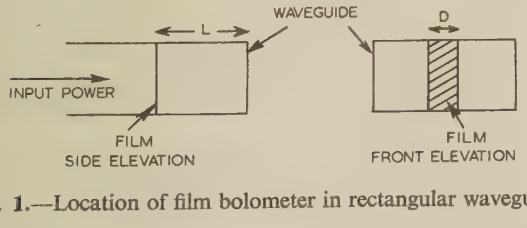


Fig. 1.—Location of film bolometer in rectangular waveguide.

nput impedance, Z, at the plane of the film can be represented y a parallel combination of the reactance, X_s, of the short- ircuited line and the equivalent circuit of the film. The latter ircuit is regarded as the effective resistance of the film, R, in eries with an inductive reactance, X_F. We have, therefore,

$$Z = \frac{(R + jX_F)jX_s}{R + j(X_F + X_s)} \quad . \quad . \quad . \quad . \quad (1)$$

uppose that, by adjustment of the film resistance and the position of the short-circuit, a non-reflecting termination is obtained. Z then becomes real and equivalent to a characteristic impedance whose numerical value is to be determined, say Z_0. It can be shown that Z is real when

$$X_s = -(R^2 + X_F^2)/X_F \quad . \quad . \quad . \quad . \quad . \quad (2)$$

When this condition is fulfilled with zero reflection coefficient,

$$Z_0 = R_m + X_F^2/R_m \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where R_m is the film resistance corresponding to the 'matched' (i.e. non-reflecting) condition.

But X_s is also given by

$$X_s = Z_0 \tan(2\pi l/\lambda_g) \quad . \quad . \quad . \quad . \quad . \quad (4)$$

and, combining eqns. (2), (3) and (4), we have

$$\tan(2\pi l/\lambda_g) = -R_m/X_F \quad . \quad . \quad . \quad . \quad (5)$$

Now l, the separation between the film and the short-circuit, can be measured experimentally. In addition, we may put R_m equal to the d.c. resistance of the metallic film, since the latter has a thickness of the order of 10^{-6} cm. The resulting value of X_F can then be substituted in eqn. (3) to give a numerical value of Z_0 in ohms. The actual value is of interest, not only in the particular problem of film-bolometer design, but also because of its significance in the theoretical definition of impedance in rectangular waveguides.

(3) IMPEDANCE MEASUREMENTS ON FILM BOLOMETERS

In the experiments to be described the films consisted of a Nichrome deposit on glass strips, 0·012 cm thick and approximately 0·4 cm in width. The Nichrome was covered with a protective layer of silicon monoxide.

(3.1) Measurements on Tunable Bolometer

In a tunable form of the instrument the waveguide beyond the film is terminated by a short-circuiting plunger of the non-contact type operated by a micrometer. The input voltage standing-wave ratio (v.s.w.r.) was measured as a function of micrometer setting at several frequencies in the 8·5–10 Gc/s band, and from these results the optimum v.s.w.r. and corresponding micrometer setting were determined for any given frequency. The results are shown in Fig. 2, which illustrates that an input v.s.w.r. very close to unity was obtained at 8·55 Gc/s. The corresponding optimum spacing between the film and the plunger was measured by a travelling microscope. The relevant data are as follows:

$$R_m = 509 \text{ ohms}$$
$$l/\lambda_g = 0{\cdot}272 \text{ (at } 8{\cdot}55 \text{ Gc/s)}$$
$$X_F = 71 \text{ ohms}$$
$$d \text{ (film width)} = 0{\cdot}397 \text{ cm}.$$

Using the above value of X_F in eqn. (3) we obtain $Z_0 = 519$ ohms. The significance of this result is discussed in Section 3.3.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
The paper is an official communication from the Radio Research Station, Department of Scientific and Industrial Research.

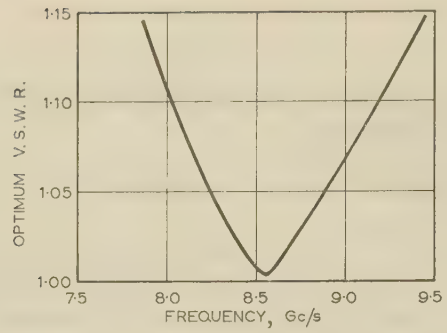


Fig. 2.—Optimum v.s.w.r. as a function of frequency for a tunable bolometer.

(3.2) Measurements on Pretuned Bolometer

A procedure somewhat different from that outlined above was followed in the measurements made on pretuned instruments (i.e. those having a fixed spacing between film and short-circuit). Several films were prepared having d.c. resistances in the range 400–500 ohms; they were then inserted in silver-plated waveguide mounts having a spacing of 1·30 cm between the film and the short-circuit. Input v.s.w.r. measurements were then made as before.

Fig. 3 shows the measurements obtained on a film 0·356 cm

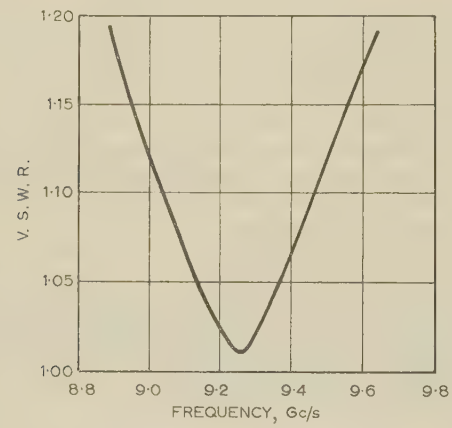


Fig. 3.—Voltage standing-wave ratio as a function of frequency for a pretuned bolometer.

wide with a d.c. resistance of 446 ohms. The optimum value of v.s.w.r. is 1·01 and occurs at a frequency of 9·26 Gc/s. (It may be mentioned here that the performance represented by Fig. 3 is a considerable improvement on that obtained in the early work on transverse-film bolometers.) For this instrument $l/\lambda_g = 0{\cdot}282$ and $X_F = 91$ ohms; thus from eqn. (3), $Z_0 = 465$ ohms.

(3.3) Discussion of Impedance Measurements

In a previous note[3] attention was drawn to the apparent relationship between the d.c. resistance of transverse-film bolometers of optimum v.s.w.r. and characteristic impedance, Z_{wv}, defined on a power/voltage basis. It is of interest to examine, in a similar manner, the more detailed results obtained above.

For a rectangular waveguide of width a and height b propagating the dominant mode, Z_{wv} is given by

$$Z_{wv} = \frac{2b}{a}\frac{377\lambda_g}{\lambda} \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Table 1 summarizes a comparision between Z_{wv} and Z_0 for the conditions described in Sections 3.1 and 3.2.

Table 1

IMPEDANCE OF BOLOMETER MOUNTS WITH TRANSVERSE FILMS

Type of instrument	Z_0	Z_{wv}
	ohms	ohms
Tunable (8·55 Gc/s) ..	519	521
Fixed plunger (9·26 Gc/s)	465	478

There is obviously good agreement between the values of Z_0 derived from the equations in Section 2 and the values of Z_{wv}. (Part of the small discrepancy between the two figures for the pretuned instrument may be due to the error in the value of Z_0 resulting from the imperfect nature of the terminating load.) Although this relationship between the characteristic impedances derived by two distinct methods may be limited in its application, the analysis would appear to be valid for the practical conditions outlined above. Some support for this view has been given by Lewin.[4] It may be noted here that the values of X_F derived above are approximately 15% lower than those calculated, with Z_{wv} as in eqn. (6), from the equation given by Marcuvitz[5] for perfectly conducting inductive strips. In the absence of any detailed theoretical treatment of a similar nature for semiconducting strips, the above results appear to yield realistic values for the reactive component of the transverse film.

The impedance measurements are of additional interest since they determine the optimum bandwidth obtainable with film bolometers of the type described. For an input v.s.w.r. of 1·5 or less a bandwidth of at least 1·2 Gc/s can be achieved with the tunable-type instrument in the 3 cm waveband.

(4) PERFORMANCE OF TUNABLE BOLOMETER MOUNT

The initial investigations with transverse-film bolometers at 3 cm indicated that the accuracy in power measurements, assuming the validity of the d.c. calibration process, was appreciably greater than that of most thermistor and bolometer milliwattmeters. These results were obtained with experimental pretuned mounts containing unprotected platinum films deposited on thin mica. Further measurements of the performance of tunable instruments of improved design are therefore of interest, and some results obtained on a commercial model of this type[6] are described below.

The film bolometer was compared against a calibrated thermistor bridge at 9·18 Gc/s using the arrangement shown in Fig. 4. The deflection of a galvanometer connected to the thermocouple terminals of the film bolometer was measured on a 50 cm scale, calibrated in milliwatts of d.c. power. Readings

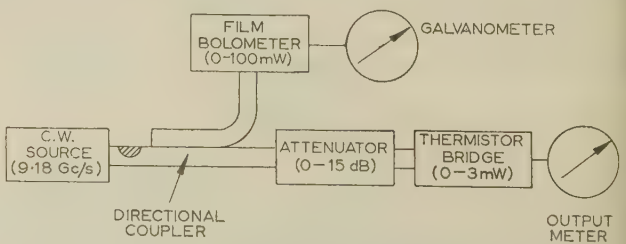
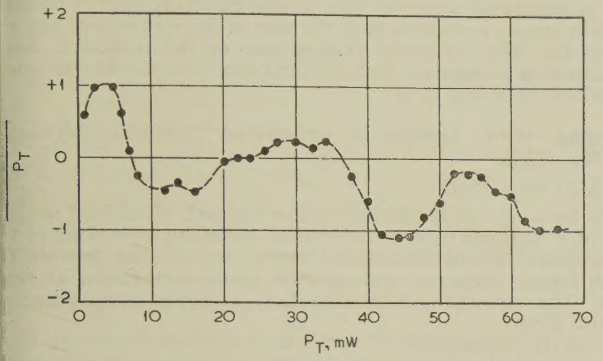


Fig. 4.—Comparison of tunable film bolometer with calibrated thermistor bridge.

this scale were then compared with those given by the rmistor bridge on the opposite arm of the directional coupler. calibrated attenuator was used with the thermistor bridge at wer levels in excess of 3 mW. Preliminary experiments were ried out in which the thermistor bridge was calibrated against orce-operated wattmeter[7] and a water calorimeter.

The results of the comparison are shown in Fig. 5, in which



g. 5.—Power measured by tunable film bolometer, P_F, compared with readings of calibrated thermistor bridge, P_T.

and P_T are the power levels derived from readings of the film lometer and calibrated thermistor respectively. In the power ige 1–70 mW the maximum difference, expressed as a per-ntage, is 1·1%. It should be emphasized, however, that there an uncertainty of about ±2% in the value of P_T, owing to effect of cumulative errors in the calibration of the bridge, enuator and directional coupler. The film bolometer investi-ed thus indicates true power within the limits of accuracy h which a comparison of the above nature can be made. ore precise power standards are required before any detailed dy of the limits of accuracy of the film bolometer can be ried out.

(5) CONCLUSIONS

The impedance measurements described in the paper, in con-nction with the derived equations, are generally applicable in e design of film-type bolometers of optimum performance. In rticular, realistic values are derived for the characteristic impedance of the rectangular waveguide and for the reactive component of the transverse film.

Measurements on a tunable bolometer mount containing a Nichrome film supported on thin glass show that an input v.s.w.r. very close to unity can be obtained at a spot frequency, with a bandwidth of about 1·2 Gc/s for a v.s.w.r. less than 1·1. The error in power measurements with such an instrument is probably not more than ±2% at levels between 1 and 100 mW in the 3 cm band. In addition, the provision of a protective layer of silicon monoxide should ensure that this performance is maintained over a long period of time, thus avoiding the necessity for frequent d.c. calibrations.

(6) ACKNOWLEDGMENTS

The authors are indebted to Mr. F. G. Clifford, of the Radio Branch of the General Post Office, for a number of valuable comments on the design of film bolometers. The work described in the paper was carried out as part of the programme of the Radio Research Board and is published by permission of the Director of Radio Research of the Department of Scientific and Industrial Research.

(7) REFERENCES

(1) LANE, J. A.: 'A Film Radiometer for Centimetre Wavelengths', *Nature*, 1956, **177**, p. 392.

(2) LANE, J. A.: 'Transverse Film Bolometers for the Measurement of Power in Rectangular Waveguides', *Proceedings I.E.E.*, Paper No. 2488 R, January, 1958 (**105** B, p. 77).

(3) LANE, J. A.: 'A Physical Interpretation of Impedance for Rectangular Waveguides', *Proceedings of the Physical Society*, 1957, **70** B, p. 1173.

(4) LEWIN, L.: 'A Physical Interpretation of Impedance for Rectangular Waveguides', *ibid.*, 1958, **71** B, p. 868.

(5) MARCUVITZ, N.: 'Waveguide Handbook', Radiation Laboratory Series (McGraw-Hill, 1917), Vol. 10, p. 227.

(6) LEMCO, I., and ROGAL, B.: 'Resistive-Film Milliwattmeters for the Frequency Bands 8·2–12·4 Gc/s, 12·4–18 Gc/s and 26·5–40 Gc/s'. *Proceedings I.E.E.*, Paper No. 3298 E, September, 1960 (**107** B, p. 427).

(7) CULLEN, A. L., ROGAL, B., and OKAMURA, S.: 'A Wide-Band Double-Vane Torque-Operated Wattmeter for 3 cm Microwaves', *Transactions of the Institute of Radio Engineers*, 1958, **MTT–6**, p. 135.

PAPER AND MONOGRAPHS PUBLISHED INDIVIDUALLY

Summaries are given below of a paper and a number of monographs which have been published individually. The paper is free of charge; the price of the monographs is 2s. each (post free). Applications, quoting the serial numbers as well as the authors' names, and accompanied by a remittance where appropriate, should be addressed to the Secretary. For convenience, books of five vouchers, price 10s., can be supplied.

Comparison of Argon, Krypton and Xenon as Admixtures in Neon Glow-Discharge Reference Tubes. Paper No. 3420 E.

F. A. BENSON, D.Eng., Ph.D., and G. P. BURDETT, B.Eng.

Special glow-discharge reference tubes containing neon-krypton, neon-xenon, neon-0·3% argon-krypton and neon-0·3% argon-xenon mixtures have been examined, the krypton or xenon content varying from 0·001 to 1%. The tubes had molybdenum electrodes and the total gas pressure was either 40 or 20 mm Hg. The striking voltage, running-voltage/current, running-voltage/temperature, initial drift, impedance/frequency and noise-voltage/current characteristics have been determined. The results of the studies are presented and discussed and are compared with those previously obtained for neon-argon tubes. It is concluded that certain desired tube characteristics can frequently be obtained by the correct choice of the gas mixture.

Stability of a Feedback System containing a Limited-Field-of-View Error Detector. Monograph No. 420 M.

D. P. ATHERTON, B.Eng.

A class of non-linear error-detector characteristics is considered where, for large magnitudes of error, the error signal decreases with increased error. Sinusoidal and random-input analysis of a closed-loop system containing such a non-linearity is shown to necessitate the calculation of a family of describing-function curves to cover operation with a d.c. input signal to the non-linear error detector. The peculiarities of such a non-linear system are discussed, and a simple method using a non-linear circuit to correct the adverse behaviour is described.

Frequency Response of Feedback Relay Amplifiers. Monograph No. 423 M.

ZE'EV BONENN, B.Sc.

The frequency response of feedback relay amplifiers is analysed by the dual-input describing-function method. At low signal frequencies the amplifier behaves like a saturating linear amplifier. This picture is modified at higher frequencies, owing to synchronous effects, and it is difficult to achieve satisfactory operation when the ratio of the input-signal frequency to the periodic excitation frequency is appreciable. To eliminate synchronous effects within the signal band, restrictions must be imposed on the forward linear-transfer function. The theory is compared with an experimental example.

Sign Matrices and Realizability of Conductance Matrices. Monograph No. 424 E.

G. BIORCI.

Given a conductance matrix G of order n, the problem of realizing it by a network with $n + 1$ nodes can be split in two parts: to find the tree of the ports corresponding to the given matrix G; and to determine the actual resistances of the branches of the network.

A procedure to solve the first (topological) problem is proposed, which is simply concerned with the sign of the elements of the given matrix G. The successful development of the procedure can be considered as a necessary and sufficient condition for the existence of a tree corresponding to G.

Travelling Wave Analysis of Generalized Networks. Monograph No. 425 E.

J. ZAWELS, Ph.D.

The behaviour of networks, whether lumped, distributed, active or passive, is analysed from a fundamental travelling-wave point of view. Generalized wave parameters are derived for a 2-port network from which image, conjugate and iterative parameters follow as special cases. The relationship of these to short- and open-circuit parameters are also given.

A matrix organization of waves in a multi-port network is next presented, which distinguishes between waves internal and external to the network. Through special transformations, voltage- and power-gain matrices are derived and their relationship to the conventional scattering matrix is shown.

Finally, a visual representation of the wave trains set up in cascaded stages is given in the form of wave flow diagrams, reminiscent of Mason's feedback graphs.

Optimum Sampled-Data Control. Monograph No. 426 M.

R. JACKSON, M.A.

A method of rendering feedback control systems amenable to treatment by the Wiener theory is applied to the case in which the controller operates on a sampled measurement. An explicit expression is obtained for the minimum attainable mean-square error for certain classes of system transfer functions and disturbance power spectra, and the form of the optimum controller is derived. The results show the inherent limitations in controllability imposed by the structure of the controlled system and by the sampling process.

The Stability of Permanent Magnets. Monograph No. 427 M.

C. E. WEBB, B.Sc.(Eng.).

The paper describes tests made to compare the magnetic stability of representative martensitic and precipitation-hardening (isotropic and anisotropic) permanent-magnet materials. The weakening of magnets with time after magnetization, when left as free from disturbance as possible, was measured over a period of about three years, both on unstabilized magnets and on magnets artificially stabilized by weakening them 1% or 5% by applying demagnetizing fields. Anisotropic alloys were found to be much more stable than isotropic alloys, the stability of Columax being outstanding. Alnico, in spite of its much higher coercive force, was not more stable than the martensitic steels.

The effect on stability of the working point of the permanent-magnet material was also examined and found to be small above the $(BH)_{max}$ point on the demagnetization curve. Below the $(BH)_{max}$ point, however, the stability of both isotropic and anisotropic materials was considerably reduced.

Further tests were made on the effects of heating at various temperatures up to 220° C and of mechanical stress and impact. Stability in high-temperature treatment is only roughly related to stability in time tests. The results of the mechanical tests were inconclusive as the magnet assemblies available, designed primarily for long-term tests, were not suitable for subjection to severe mechanical treatment.

PROCEEDINGS OF THE INSTITUTION OF ELECTRICAL ENGINEERS

Part B. ELECTRONIC AND COMMUNICATION ENGINEERING (INCLUDING RADIO ENGINEERING), JANUARY 1961

CONTENTS

Bibliographical References.—It is requested that bibliographical reference to an Institution paper should always include the serial number of the paper and the month and year of publication, which will be found at the top right-hand corner of the first page of the paper. This information should precede the reference to the Volume and Part.

Example.—SMITH, J.: 'Reflections from the Ionosphere', *Proceedings I.E.E.*, Paper No. 5001 E, December, 1960 (107 B, p. 1234).

Published by The Institution, Savoy Place, London, W.C.2. *Telephone:* COVent Garden 1871. *Telegrams:* 'Voltampere, Phone, London.
Printed by Unwin Brothers Limited, Woking and London.